

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



29

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification⁶ : C12N 5/14, 15/00, 15/05, 15/09, 15/29, 15/31, 15/64, 15/82, A01G 13/00, A01H 1/00</p>	<p>A1</p>	<p>(11) International Publication Number: WO 97/17432 (43) International Publication Date: 15 May 1997 (15.05.97)</p>
<p>(21) International Application Number: PCT/US96/18003 (22) International Filing Date: 6 November 1996 (06.11.96) (30) Priority Data: 60/007,255 6 November 1995 (06.11.95) US 08/608,423 28 February 1996 (28.02.96) US 08/705,484 28 August 1996 (28.08.96) US (71) Applicant: WISCONSIN ALUMNI RESEARCH FOUNDATION [US/US]; 614 North Walnut Street, Madison, WI 53707-7365 (US).</p>		<p>(72) Inventors: ENSIGN, Jerald, C.; 1810 North Walnut Street, Madison, WI 53705 (US). BOWEN, David, J.; 5668 Highway A, Oregon, WI 53575 (US). PETELL, James; 1427 Hunters Glen, Zionsville, IN 46077 (US). FATIG, Raymond; 30 Clay Court, Zionsville, IN 46077 (US). SCHOONOVER, Sue; 7142 Marstella, Brownsburg, IN 46112 (US). FFRENCH-CONSTANT, Richard, H.; 1006 University Bay Drive, Madison, WI 53705 (US). ROCHE-LEAU, Thomas, A.; 3100 Buena Vista Street, Madison, WI 53704 (US). BLACKBURN, Michael, B.; 2127 Luann Lane, Madison, WI 53713 (US). HEY, Timothy, D.; 1653 Catalina Way, Zionsville, IN 46077 (US). MERLO, Donald, J.; 11845 Durbin Drive, Carmel, IN 46032 (US). ORR, Gregory, L.; 1028 Saratoga Circle, Indianapolis, IN 46280 (US). ROBERTS, Jean, L.; 26035 State Road 19, Arcadia, IN 46030 (US). STRICKLAND, James, A.; 780 Mt. Zion Road, Lebanon, IN 46052 (US). GUO, Lining; 7 Nelson Circle, Brownsburg, IN 46112 (US). CICHE, Todd, A.; 1609 Chadbourne Avenue, Madison, WI 53705 (US). (74) Agent: SEAY, Nicholas, J.; Quarles & Brady, P.O. Box 2113, Madison, WI 53701-2113 (US). (81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published With international search report.</p>
<p>(54) Title: INSECTICIDAL PROTEIN TOXINS FROM PHOTORHABDUS (57) Abstract Proteins from the genus <i>Photorhabdus</i> are toxic to insects upon exposure. <i>Photorhabdus luminescens</i> (formerly <i>Xenorhabdus luminescens</i>) have been found in mammalian clinical samples and as a bacterial symbiont of entomopathogenic nematodes of genus <i>Heterorhabditis</i>. These protein toxins can be applied to, or genetically engineered into, insect larvae food and plants for insect control.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Larvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

INSECTICIDAL PROTEIN TOXINS FROM *PHOTORHABDUS*

5

10

15

Field of the Invention

The present invention relates to toxins isolated from bacteria and the use of said toxins as insecticides.

20

Background of the Invention

Many insects are widely regarded as pests to homeowners, to picnickers, to gardeners, and to farmers and others whose investments in agricultural products are often destroyed or diminished as a result of insect damage to field crops. Particularly in areas where the growing season is short, significant insect damage can mean the loss of all profits to growers and a dramatic decrease in crop yield. Scarce supply of particular agricultural products invariably results in higher costs to food processors and, then, to the ultimate consumers of food plants and products derived from those plants.

Preventing insect damage to crops and flowers and eliminating the nuisance of insect pests have typically relied on strong organic pesticides and insecticides with broad toxicities. These synthetic products have come under attack by the general population as being too harsh on the environment and on those exposed to such agents. Similarly in non-agricultural settings, homeowners would be satisfied to have insects avoid their homes or outdoor meals without needing to kill the insects.

The extensive use of chemical insecticides has raised environmental and health concerns for farmers, companies that produce the insecticides, government agencies, public interest groups, and the public in general. The development of less
5 intrusive pest management strategies has been spurred along both by societal concern for the environment and by the development of biological tools which exploit mechanisms of insect management. Biological control agents present a promising alternative to chemical insecticides.

10 Organisms at every evolutionary development level have devised means to enhance their own success and survival. The use of biological molecules as tools of defense and aggression is known throughout the animal and plant kingdoms. In addition, the relatively new tools of the genetic engineer allow modifications
15 to biological insecticides to accomplish particular solutions to particular problems.

One such agent, *Bacillus thuringiensis* (Bt), is an effective insecticidal agent, and is widely commercially used as such. In fact, the insecticidal agent of the Bt bacterium is a protein
20 which has such limited toxicity, it can be used on human food crops on the day of harvest. To non-targeted organisms, the Bt toxin is a digestible non-toxic protein.

Another known class of biological insect control agents are certain genera of nematodes known to be vectors of transmission
25 for insect-killing bacterial symbionts. Nematodes containing insecticidal bacteria invade insect larvae. The bacteria then kill the larvae. The nematodes reproduce in the larval cadaver. The nematode progeny then eat the cadaver from within. The bacteria-containing nematode progeny thus produced can then
30 invade additional larvae.

In the past, insecticidal nematodes in the *Steinernema* and *Heterorhabditis* genera were used as insect control agents. Apparently, each genus of nematode hosts a particular species of bacterium. In nematodes of the *Heterorhabditis* genus, the
35 symbiotic bacterium is *Photorhabdus luminescens*.

Although these nematodes are effective insect control agents, it is presently difficult, expensive, and inefficient to produce, maintain, and distribute nematodes for insect control.

It has been known in the art that one may isolate an
40 insecticidal toxin from *Photorhabdus luminescens* that has

activity only when injected into Lepidopteran and Coleopteran insect larvae. This has made it impossible to effectively exploit the insecticidal properties of the nematode or its bacterial symbiont. What would be useful would be a more
5 practical, less labor-intensive wide-area delivery method of an insecticidal toxin which would retain its biological properties after delivery. It would be quite desirous to discover toxins with oral activity produced by the genus *Photorhabdus*. The isolation and use of these toxins are desirous due to efficacious
10 reasons. Until applicants' discoveries, these toxins had not been isolated or characterized.

Summary of the Invention

15 The native toxins are protein complexes that are produced and secreted by growing bacteria cells of the genus *Photorhabdus*, of interest are the proteins produced by the species *Photorhabdus luminescens*. The protein complexes, with a molecular size of approximately 1,000 kDa, can be separated by SDS-PAGE gel
20 analysis into numerous component proteins. The toxins contain no hemolysin, lipase, type C phospholipase, or nuclease activities. The toxins exhibit significant toxicity upon exposure administration to a number of insects.

The present invention provides an easily administered
25 insecticidal protein as well as the expression of toxin in a heterologous system.

The present invention also provides a method for delivering insecticidal toxins that are functional active and effective against many orders of insects.

30 Objects, advantages, and features of the present invention will become apparent from the following specification.

Brief Description of the Drawings

35 Fig. 1 is an illustration of a match of cloned DNA isolates used as a part of sequence genes for the toxin of the present invention.

Fig. 2 is a map of three plasmids used in the sequencing process.

Fig. 3 is a map illustrating the inter-relationship of several partial DNA fragments.

Fig. 4 is an illustration of a homology analysis between the protein sequences of TcbAii and TcaBii proteins.

5 Fig. 5 is a phenogram of *Photorhabdus* strains. Relationship of *Photorhabdus* Strains was defined by rep-PCR. The upper axis of Fig. 5 measures the percentage similarity of strains based on scoring of rep-PCR products (i.e., 0.0 [no similarity] to 1.0 [100% similarity]). At the right axis, the numbers and letters indicate the various strains tested; 14=W-14, Hm=Hm, H9=H9, 7=WX-7, 1=WX-1, 2=WX-2, 88=HP88, NC-1=NC-1, 4=WX-4, 10 9=WX-9, 8=WX-8, 10=WX-10, WIR=WIR, 3=WX-3, 11=WX-11, 5=WX-5, 6=WX-6, 12=WX-12, x14=WX-14, 15=WX-15, Hb=Hb, B2=B2, 48 through 52=ATCC 43948 through ATCC 43952. Vertical lines separating horizontal lines indicate the degree of relatedness (as read from the extrapolated intersection of the vertical line with the upper axis) between strains or groups of strains at the base of the horizontal lines (e.g., strain W-14 is approximately 60% similar to strains H9 and Hm).

20 Fig. 6 is an illustration of the genomic maps of the W-14 Strain.

Detailed Description of the Invention

25 The present inventions are directed to the discovery of a unique class of insecticidal protein toxins from the genus *Photorhabdus* that have oral toxicity against insects. A unique feature of *Photorhabdus* is its bioluminescence. *Photorhabdus* may be isolated from a variety of sources. One such source is 30 nematodes, more particularly nematodes of the genus *Heterorhabditis*. Another such source is from human clinical samples from wounds, see Farmer et al. 1989 J. Clin. Microbiol. 27 pp. 1594-1600. These saprophytic strains are deposited in the American Type Culture Collection (Rockville, MD) ATCC #s 43948, 35 43949, 43950, 43951, and 43952, and are incorporated herein by reference. It is possible that other sources could harbor *Photorhabdus* bacteria that produce insecticidal toxins. Such sources in the environment could be either terrestrial or aquatic based.

The genus *Photorhabdus* is taxonomically defined as a member of the Family *Enterobacteriaceae*, although it has certain traits atypical of this family. For example, strains of this genus are nitrate reduction negative, yellow and red pigment producing and bioluminescent. This latter trait is otherwise unknown within the *Enterobacteriaceae*. *Photorhabdus* has only recently been described as a genus separate from the *Xenorhabdus* (Boemare et al., 1993 Int. J. Syst. Bacteriol. 43, 249-255). This differentiation is based on DNA-DNA hybridization studies, phenotypic differences (e.g., presence (*Photorhabdus*) or absence (*Xenorhabdus*) of catalase and bioluminescence) and the Family of the nematode host (*Xenorhabdus*; *Steinernematidae*, *Photorhabdus*; *Heterorhabditidae*). Comparative, cellular fatty-acid analyses (Janse et al. 1990, Lett. Appl. Microbiol 10, 131-135; Suzuki et al. 1990, J. Gen. Appl. Microbiol., 36, 393-401) support the separation of *Photorhabdus* from *Xenorhabdus*.

In order to establish that the strain collection disclosed herein was comprised of *Photorhabdus* strains, the strains were characterized based on recognized traits which define *Photorhabdus* and differentiate it from other *Enterobacteriaceae* and *Xenorhabdus* species. (Farmer, 1984 Bergey's Manual of Systemic Bacteriology Vol. 1 pp.510-511; Akhurst and Boemare 1988, J. Gen. Microbiol. 134 pp.1835-1845; Boemare et al. 1993 Int. J. Syst. Bacteriol. 43 pp.249-255, which are incorporated herein by reference). The traits studied were the following: gram stain negative rods, organism size, colony pigmentation, inclusion bodies, presence of catalase, ability to reduce nitrate, bioluminescence, dye uptake, gelatin hydrolysis, growth on selective media, growth temperature, survival under anerobic conditions and motility. Fatty acid analysis was used to confirm that the strains herein all belong to the single genus *Photorhabdus*.

Currently, the bacterial genus *Photorhabdus* is comprised of a single defined species, *Photorhabdus luminescens* (ATCC Type strain #29999, Poinar et al., 1977, Nematologica 23, 97-102). A variety of related strains have been described in the literature (e.g. Akhurst et al. 1988 J. Gen. Microbiol., 134, 1835-1845; Boemare et al. 1993 Int. J. Syst. Bacteriol. 43 pp. 249-255; Putz et al. 1990, Appl. Environ. Microbiol., 56, 181-186). Numerous

Photorhabdus strains have been characterized herein. Such strains are listed in Table 18 in the Examples. Because there is currently only one species (*luminescens*) defined within the genus *Photorhabdus*, the *luminescens* species traits were used to

5 characterize the strains herein. As can be seen in Fig. 5, these strains are quite diverse. It is not unforeseen that in the future there may be other *Photorhabdus* species that will have some of the attributes of the *luminescens* species as well as some different characteristics that are presently not defined as a

10 trait of *Photorhabdus luminescens*. However, the scope of the invention herein is to any *Photorhabdus* species or strains which produce proteins that have functional activity as insect control agents, regardless of other traits and characteristics.

Furthermore, as is demonstrated herein, the bacteria of the

15 genus *Photorhabdus* produce proteins that have functional activity as defined herein. Of particular interest are proteins produced by the species *Photorhabdus luminescens*. The inventions herein should in no way be limited to the strains which are disclosed herein. These strains illustrate for the first time that

20 proteins produced by diverse isolates of *Photorhabdus* are toxic upon exposure to insects. Thus, included within the inventions described herein are the strains specified herein and any mutants thereof, as well as any strains or species of the genus

Photorhabdus that have the functional activity described herein.

25 There are several terms that are used herein that have a particular meaning and are as follows:

By "functional activity" it is meant herein that the protein toxins function as insect control agents in that the proteins are

30 orally active, or have a toxic effect, or are able to disrupt or deter feeding, which may or may not cause death of the insect. When an insect comes into contact with an effective amount of toxin delivered via transgenic plant expression, formulated protein compositions(s), sprayable protein composition(s), a bait

35 matrix or other delivery system, the results are typically death of the insect, or the insects do not feed upon the source which makes the toxins available to the insects.

The protein toxins discussed herein are typically referred to as "insecticides". By insecticides it is meant herein that the protein toxins have a "functional activity" as further defined herein and are used as insect control agents.

5

By the use of the term "oligonucleotides" it is meant a macromolecule consisting of a short chain of nucleotides of either RNA or DNA. Such length could be at least one nucleotide, but typically are in the range of about 10 to about 12

10 nucleotides. The determination of the length of the oligonucleotide is well within the skill of an artisan and should not be a limitation herein. Therefore, oligonucleotides may be less than 10 or greater than 12.

15 By the use of the term "toxic" or "toxicity" as used herein it is meant that the toxins produced by *Photorhabdus* have "functional activity" as defined herein.

By the use of the term "genetic material" herein, it is meant to
20 include all genes, nucleic acid, DNA and RNA.

Fermentation broths from selected strains reported in Table 18 were used to determine the following: breadth of insecticidal toxin production by the *Photorhabdus* genus, the
25 insecticidal spectrum of these toxins, and to provide source material to purify the toxin complexes. The strains characterized herein have been shown to have oral toxicity against a variety of insect orders. Such insect orders include but are not limited to *Coleoptera*, *Homoptera*, *Lepidoptera*,
30 *Diptera*, *Acarina*, *Hymenoptera* and *Dictyoptera*.

As with other bacterial toxins, the rate of mutation of the bacteria in a population causes many related toxins slightly different in sequence to exist. Toxins of interest here are
35 those which produce protein complexes toxic to a variety of insects upon exposure, as described herein. Preferably, the toxins are active against *Lepidoptera*, *Coleoptera*, *Homopotera*, *Diptera*, *Hymenoptera*, *Dictyoptera* and *Acarina*. The inventions herein are intended to capture the protein toxins homologous to
40 protein toxins produced by the strains herein and any derivative

5 By the use of the term "*Photorhabdus* toxin" it is meant any protein produced by a *Photorhabdus* microorganism strain which has functional activity against insects, where the *Photorhabdus* toxin could be formulated as a sprayable composition, expressed by a transgenic plant, formulated as
10 a bait matrix, delivered via a Baculovirus, or delivered by any other applicable host or delivery system.

strains thereof, as well as any protein toxins produced by *Photorhabdus*. These homologous proteins may differ in sequence, but do not differ in function from those toxins described herein. Homologous toxins are meant to include protein complexes of
5 between 300 kDa to 2,000 kDa and are comprised of at least two
(2) subunits, where a subunit is a peptide which may or may not be the same as the other subunit. Various protein subunits have been identified and are taught in the Examples herein. Typically, the protein subunits are between about 18 kDa to about
10 230 kDa; between about 160 kDa to about 230 kDa; 100 kDa to 160 kDa; about 80 kDa to about 100 kDa; and about 50 kDa to about 80 kDa.

As discussed above, some *Photorhabdus* strains can be isolated from nematodes. Some nematodes, elongated cylindrical
15 parasitic worms of the phylum *Nematoda*, have evolved an ability to exploit insect larvae as a favored growth environment. The insect larvae provide a source of food for growing nematodes and an environment in which to reproduce. One dramatic effect that follows invasion of larvae by certain nematodes is larval death.
20 Larval death results from the presence of, in certain nematodes, bacteria that produce an insecticidal toxin which arrests larval growth and inhibits feeding activity.

Interestingly, it appears that each genus of insect parasitic nematode hosts a particular species of bacterium,
25 uniquely adapted for symbiotic growth with that nematode. In the interim since this research was initiated, the name of the bacterial genus *Xenorhabdus* was reclassified into the *Xenorhabdus* and the *Photorhabdus*. Bacteria of the genus *Photorhabdus* are characterized as being symbionts of *Heterorhabditus* nematodes
30 while *Xenorhabdus* species are symbionts of the *Steinernema* species. This change in nomenclature is reflected in this specification, but in no way should a change in nomenclature alter the scope of the inventions described herein.

The peptides and genes that are disclosed herein are named
35 according to the guidelines recently published in the Journal of Bacteriology "Instructions to Authors" p. i-xii (Jan. 1996), which is incorporated herein by reference. The following peptides and genes were isolated from *Photorhabdus* strain W-14.

Peptide / Gene Nomenclature
Toxin complex (Tc)

	Peptide Name	Gene Name	Patent Sequence ID#
5	<u>tca genomic region</u>		
	TcaA	tcaA	12
	TcaAiii	tcaA	4
10	TcaBi	tcaB	3 (19, 20)
	TcaBii	tcaB	5
	TcaC	tcaC	2
	<u>tcb genomic region</u>		
15	TcbA	tcbA	16
	TcbAi	tcbA	(pro-peptide)
	TcbAii	tcbA	1 (21, 22, 23, 24)
	TcbAiii	tcbA	40
20	<u>tcc genomic region</u>		
	TccA	tccA	8
	TccB	tccB	7
	<u>tcd genomic region</u>		
25	TcdAi	tcdA	(pro-peptide)
	TcdAii	tcdA	13, (38, 39 17, 18)
	TcdAiii	tcdA	41, (42, 43)
30	TcdB	tcdB	14

(bracket sequence indicates internal amino acid sequence obtained by tryptic digests)

35 The sequences listed above are grouped by genomic region. The *tcbA* gene was expressed in *E. coli* as two protein fragments TcbA and TcbAiii as illustrated in the Examples. It may be beneficial to have proteolytic clippage of some sequences to obtain the higher activity of the toxins for commercial
40 transgenic applications.

 The toxins described herein are quite unique in that the toxins have functional activity, which is key to developing an insect management strategy. In developing an insect management
45 strategy, it is possible to delay or circumvent the protein degradation process by injecting a protein directly into an organism, avoiding its digestive tract. In such cases, the protein administered to the organism will retain its function until it is denatured, non-specifically degraded, or eliminated
50 by the immune system in higher organisms. Injection into insects

of an insecticidal toxin has potential application only in the laboratory, and then only on large insects which are easily injected. The observation that the insecticidal protein toxins herein described exhibits their toxic activity after oral
5 ingestion or contact with the toxins permits the development of an insect management plan based solely on the ability to incorporate the protein toxins into the insect diet. Such a plan could result in the production of insect baits.

The *Photorhabdus* toxins may be administered to insects in a
10 purified form. The toxins may also be delivered in amounts from about 1 to about 100 mg / liter of broth. This may vary upon formulation condition, conditions of the inoculum source, techniques for isolation of the toxin, and the like. The toxins may be administered as an exudate secretion or cellular protein
15 originally expressed in a heterologous prokaryotic or eukaryotic host. Bacteria are typically the hosts in which proteins are expressed. Eukaryotic hosts could include but are not limited to plants, insects and yeast. Alternatively, the toxins may be produced in bacteria or transgenic plants in the field or in the
20 insect by a baculovirus vector. Typically the toxins will be introduced to the insect by incorporating one or more of the toxins into the insects' feed.

Complete lethality to feeding insects is useful but is not required to achieve useful toxicity. If the insects avoid the
25 toxin or cease feeding, that avoidance will be useful in some applications, even if the effects are sublethal. For example, if insect resistant transgenic crop plants are desired, a reluctance of insects to feed on the plants is as useful as lethal toxicity to the insects since the ultimate objective is protection of the
30 plants rather than killing the insect.

There are many other ways in which toxins can be incorporated into an insect's diet. As an example, it is possible to adulterate the larval food source with the toxic protein by spraying the food with a protein solution, as
35 disclosed herein. Alternatively, the purified protein could be genetically engineered into an otherwise harmless bacterium, which could then be grown in culture, and either applied to the food source or allowed to reside in the soil in an area in which insect eradication was desirable. Also, the protein could be
40 genetically engineered directly into an insect food source. For

instance, the major food source of many insect larvae is plant material.

By incorporating genetic material that encodes the insecticidal properties of the *Photographus* toxins into the genome of a plant eaten by a particular insect pest, the adult or larvae would die after consuming the food plant. Numerous members of the monocotyledonous and dictyledonous genera have been transformed. Transgenic agronomic crops as well as fruits and vegetables are of commercial interest. Such crops include alfalfa, sorghum, wheat, cotton, peanuts, tomatoes, potatoes, and the like. Several techniques exist for introducing foreign genetic material into plant cells, and for obtaining plants that stably maintain and express the introduced gene. Such techniques include acceleration directly into cells (U.S. Patents 4,945,050 to Cornell and 5,141,131 to DowElanco). Plants may be transformed using *Agrobacterium* technology, see U.S. Patent 5,177,010 to University of Toledo, 5,104,310 to Texas A&M, European Patent 159418B1 and 176,112 to Schilperoot, U.S. Patents 120516, 5,469,976, 5,464,763 and 4,940,838 and 4,693,976 to Schilperoot, European Patent Applications 0267159, and 0292435 MaxPlanck, European Patent Applications 604662 and 627752 to Japan Tobacco, European Patent Applications 5,231,019 all to Ciba Geigy, U.S. Patents 5,463,174 and 4,762,785 both to Calgene, and U.S. Patents 5,004,863 and 5,159,135 both to Zeneca. Electroporation technology includes whiskers technology, see U.S. Patents 5,302,523 and 5,464,765 both to Zeneca. Other transformation technology has also been used to transform plants, see WO 87/06614 to Boyce Thompson Institute, 5,472,869 and 5,384,253 both to Dekalb, WO9209696 and WO9321335 both to PGS. All of these transformation patents and publications are incorporated by reference. In addition to numerous technologies for transforming plants, the type of tissue which is contacted with the foreign genes may vary as well. Such tissue would include but would not be limited to embryonic tissue, callus tissue type I and II, hypocotyl, meristem, and the like. Almost all plant tissues may

be transformed during dedifferentiation using appropriate techniques within the skill of an artisan.

Another variable is the choice of a selectable marker. The preference for a particular marker is at the discretion of the artisan, but any of the following selectable markers may be used along with any other gene not listed herein which could function as a selectable marker. Such selectable markers include but are not limited to aminoglycoside phosphotransferase gene of transposon Tn5 (Aph II) which encodes resistance to the antibiotics kanamycin, neomycin and G418, as well as those genes which code for resistance or tolerance to glyphosate; hygromycin; methotrexate; phosphinothricin (bialophos); imidazolinones, sulfonylureas and triazolopyrimidine herbicides, such as chlorosulfuron; bromoxynil, dalapon and the like.

In addition to a selectable marker, it may be desirable to use a reporter gene. In some instances a reporter gene may be used without a selectable marker. Reporter genes are genes which are typically not present or expressed in the recipient organism or tissue. The reporter gene typically encodes for a protein which provides for some phenotypic change or enzymatic property. Examples of such genes are provided in K. Weising et al. Ann. Rev. Genetics, 22, 421 (1988), which is incorporated herein by reference. A preferred reporter gene is the glucuronidase (GUS) gene.

Regardless of transformation technique, the gene is preferably incorporated into a gene transfer vector adapted to express the *Photobacterium* toxins in the plant cell by including in the vector a plant promoter. In addition to plant promoters, promoters from a variety of sources can be used efficiently in plant cells to express foreign genes. For example, promoters of bacterial origin, such as the octopine synthase promoter, the nopaline synthase promoter, the mannopine synthase promoter; promoters of viral origin, such as the cauliflower mosaic virus (35S and 19S) and the like may be used. Plant promoters include, but are not limited to ribulose-1,6-bisphosphate (RUBP) carboxylase small subunit (ssu), beta-conglycinin promoter, phaseolin promoter, ADH promoter, heat-shock promoters and tissue specific promoters. Promoters may also contain certain enhancer sequence elements that may improve the transcription efficiency. Typical enhancers include but are not limited to Adh-intron 1 and

Adh-intron 6. Constitutive promoters may be used. Constitutive promoters direct continuous gene expression in all cells types and at all times (e.g., actin, ubiquitin, CaMV 35S). Tissue specific promoters are responsible for gene expression in
5 specific cell or tissue types, such as the leaves or seeds (e.g., zein, oleosin, napin, ACP) and these promoters may also be used. Promoters may also be active during a certain stage of the plants' development as well as active in plant tissues and organs. Examples of such promoters include but are not limited
10 to pollen-specific, embryo specific, corn silk specific, cotton fiber specific, root specific, seed endosperm specific promoters and the like.

Under certain circumstances it may be desirable to use an inducible promoter. An inducible promoter is responsible for
15 expression of genes in response to a specific signal, such as: physical stimulus (heat shock genes); light (RUBP carboxylase); hormone (Em); metabolites; and stress. Other desirable transcription and translation elements that function in plants may be used. Numerous plant-specific gene transfer vectors are
20 known to the art.

In addition, it is known that to obtain high expression of bacterial genes in plants it is preferred to reengineer the bacterial genes so that they are more efficiently expressed in the cytoplasm of plants. Maize is one such plant where it is
25 preferred to reengineer the bacterial gene(s) prior to transformation to increase the expression level of the toxin in the plant. One reason for the reengineering is the very low G+C content of the native bacterial gene(s) (and consequent skewing towards high A+T content). This results in the generation of
30 sequences mimicking or duplicating plant gene control sequences that are known to be highly A+T rich. The presence of some A+T-rich sequences within the DNA of the gene(s) introduced into plants (e.g., TATA box regions normally found in gene promoters) may result in aberrant transcription of the gene(s). On the
35 other hand, the presence of other regulatory sequences residing in the transcribed mRNA (e.g., polyadenylation signal sequences (AAUAAA), or sequences complementary to small nuclear RNAs involved in pre-mRNA splicing) may lead to RNA instability. Therefore, one goal in the design of reengineered bacterial

gene(s), more preferably referred to as plant optimized gene(s), is to generate a DNA sequence having a higher G+C content, and preferably one close to that of plant genes coding for metabolic enzymes. Another goal in the design of the plant optimized gene(s) is to generate a DNA sequence that not only has a higher G+C content, but by modifying the sequence changes, should be made so as to not hinder translation.

An example of a plant that has a high G+C content is maize. The table below illustrates how high the G+C content is in maize. As in maize, it is thought that G+C content in other plants is also high.

Table 1
Compilation of G+C contents of protein coding regions
of maize genes

Protein Class ^a	Range %G+C	Mean %G+C ^b
Metabolic Enzymes (40)	44.4-75.3	59.0 (8.0)
Storage Proteins		
Group I (23)	46.0-51.9	48.1 (1.3)
Group II (13)	60.4-74.3	67.5 (3.2)
Group I + II (36)	46.0-74.3	55.1 (9.6) ^c
Structural Proteins (18)	48.6-70.5	63.6 (6.7)
Regulatory Proteins (5)	57.2-68.9	62.0 (4.9)
Uncharacterized Proteins (9)	41.5-70.3	64.3 (7.2)
All Proteins (108)	44.4-75.3	60.8 (5.2)

^a Number of genes in class given in parentheses.

^b Standard deviations given in parentheses.

^c Combined groups mean ignored in calculation of overall mean.

For the data in Table 1, coding regions of the genes were extracted from GenBank (Release 71) entries, and base compositions were calculated using the MacVector™ program (IBI, New Haven, CT). Intron sequences were ignored in the

calculations. Group I and II storage protein gene sequences were distinguished by their marked difference in base composition.

Due to the plasticity afforded by the redundancy of the genetic code (i.e., some amino acids are specified by more than
5 one codon), evolution of the genomes of different organisms or classes or organisms has resulted in differential usage of redundant codons. This "codon bias" is reflected in the mean base composition of protein coding regions. For example, organisms with relatively low G+C contents utilize codons having A or T in
10 the third position of redundant codons, whereas those having higher G+C contents utilize codons having G or C in the third position. It is thought that the presence of "minor" codons within a gene's mRNA may reduce the absolute translation rate of that mRNA, especially when the relative abundance of the charged
15 tRNA corresponding to the minor codon is low. An extension of this is that the diminution of translation rate by individual minor codons would be at least additive for multiple minor codons. Therefore, mRNAs having high relative contents of minor codons would have correspondingly low translation rates. This
20 rate would be reflected by the synthesis of low levels of the encoded protein.

In order to reengineer the bacterial gene(s), the codon bias of the plant is determined. The codon bias is the statistical codon distribution that the plant uses for coding its proteins.
25 After determining the bias, the percent frequency of the codons in the gene(s) of interest is determined. The primary codons preferred by the plant should be determined as well as the second and third choice of preferred codons. The amino acid sequence of the protein of interest is reverse translated so that the
30 resulting nucleic acid sequence codes for the same protein as the native bacterial gene, but the resulting nucleic acid sequence corresponds to the first preferred codons of the desired plant. The new sequence is analyzed for restriction enzyme sites that might have been created by the modification. The identified
35 sites are further modified by replacing the codons with second or third choice preferred codons. Other sites in the sequence which could affect the transcription or translation of the gene of interest are the exon:intron 5' or 3' junctions, poly A addition signals, or RNA polymerase termination signals. The sequence is

further analyzed and modified to reduce the frequency of TA or GC doublets. In addition to the doublets, G or C sequence blocks that have more than about four residues that are the same can affect transcription of the sequence. Therefore, these blocks
5 are also modified by replacing the codons of first or second choice, etc. with the next preferred codon of choice. It is preferred that the plant optimized gene(s) contains about 63% of first choice codons, between about 22% to about 37% second choice codons, and between 15% and 0% third choice codons, wherein the
10 total percentage is 100%. Most preferred the plant optimized gene(s) contain about 63% of first choice codons, at least about 22% second choice codons, about 7.5% third choice codons, and about 7.5% fourth choice codons, wherein the total percentage is 100%. The method described above enables one skilled in the art
15 to modify gene(s) that are foreign to a particular plant so that the genes are optimally expressed in plants. The method is further illustrated in pending provisional application U.S. 60/005,405 filed on October 13, 1995, which is incorporated herein by reference.

20 Thus, in order to design plant optimized gene(s) the amino acid sequence of the toxins are reverse translated into a DNA sequence, utilizing a nonredundant genetic code established from a codon bias table compiled for the gene DNA sequence for the particular plant being transformed. The resulting DNA sequence,
25 which is completely homogeneous in codon usage, is further modified to establish a DNA sequence that, besides having a higher degree of codon diversity, also contains strategically placed restriction enzyme recognition sites, desirable base composition, and a lack of sequences that might interfere with
30 transcription of the gene, or translation of the product mRNA.

It is theorized that bacterial genes may be more easily expressed in plants if the bacterial genes are expressed in the plastids. Thus, it may be possible to express bacterial genes in plants, without optimizing the genes for plant expression, and
35 obtain high express of the protein. See U.S. Patent Nos. 4,762,785; 5,451,513 and 5,545,817, which are incorporated herein by reference.

One of the issues regarding commercial exploiting transgenic plants is resistance management. This is of particular concern with *Bacillus thuringiensis* toxins. There are numerous companies commercially exploiting *Bacillus thuringiensis* and there has been
5 much concern about *Bt* toxins becoming resistant. One strategy for insect resistant management would be to combine the toxins produced by *Photobacterium* with toxins such as *Bt*, vegetative insect proteins (Ciba Geigy) or other toxins. The combinations could be formulated for a sprayable application or could be
10 molecular combinations. Plants could be transformed with *Photobacterium* genes that produce insect toxins and other insect toxin genes such as *Bt* as with other insect toxin genes such as *Bt*.

European Patent Application 0400246A1 describes
15 transformation of 2 *Bt* in a plant, which could be any 2 genes. Another way to produce a transgenic plant that contains more than one insect resistant gene would be to produce two plants, with each plant containing an insect resistant gene. These plants would be backcrossed using traditional plant breeding techniques
20 to produce a plant containing more than one insect resistant gene.

In addition to producing a transformed plant containing plant optimized gene(s), there are other delivery systems where it may be desirable to reengineer the bacterial gene(s). Along
25 the same lines, a genetically engineered, easily isolated protein toxin fusing together both a molecule attractive to insects as a food source and the insecticidal activity of the toxin may be engineered and expressed in bacteria or in eukaryotic cells using standard, well-known techniques. After purification in the
30 laboratory such a toxic agent with "built-in" bait could be packaged inside standard insect trap housings.

Another delivery scheme is the incorporation of the genetic material of toxins into a baculovirus vector. Baculoviruses infect particular insect hosts, including those desirably
35 targeted with the *Photobacterium* toxins. Infectious baculovirus harboring an expression construct for the *Photobacterium* toxins could be introduced into areas of insect infestation to thereby intoxicate or poison infected insects.

Transfer of the insecticidal properties requires nucleic acid sequences encoding the coding the amino acid sequences for the *Photobacterium* toxins integrated into a protein expression vector appropriate to the host in which the vector will reside.

5 One way to obtain a nucleic acid sequence encoding a protein with insecticidal properties is to isolate the native genetic material which produces the toxins from *Photobacterium*, using information deduced from the toxin's amino acid sequence, large portions of which are set forth below. As described below, methods of

10 purifying the proteins responsible for toxin activity are also disclosed.

Using N-terminal amino acid sequence data, such as set forth below, one can construct oligonucleotides complementary to all, or a section of, the DNA bases that encode the first amino acids

15 of the toxin. These oligonucleotides can be radiolabeled and used as molecular probes to isolate the genetic material from a genomic genetic library built from genetic material isolated from strains of *Photobacterium*. The genetic library can be cloned in plasmid, cosmid, phage or phagemid vectors. The library could be

20 transformed into *Escherichia coli* and screened for toxin production by the transformed cells using antibodies raised against the toxin or direct assays for insect toxicity.

This approach requires the production of a battery of oligonucleotides, since the degenerate genetic code allows an

25 amino acid to be encoded in the DNA by any of several three-nucleotide combinations. For example, the amino acid arginine can be encoded by nucleic acid triplets CGA, CGC, CGG, CGT, AGA, and AGG. Since one cannot predict which triplet is used at those positions in the toxin gene, one must prepare oligonucleotides

30 with each potential triplet represented. More than one DNA molecule corresponding to a protein subunit may be necessary to construct a sufficient number of oligonucleotide probes to recover all of the protein subunits necessary to achieve oral toxicity.

35 From the amino acid sequence of the purified protein, genetic materials responsible for the production of toxins can readily be isolated and cloned, in whole or in part, into an expression vector using any of several techniques well-known to one skilled in the art of molecular biology. A typical

40 expression vector is a DNA plasmid, though other transfer means

including, but not limited to, cosmids, phagemids and phage are also envisioned. In addition to features required or desired for plasmid replication, such as an origin of replication and antibiotic resistance or other form of a selectable marker such as the *bar* gene of *Streptomyces hygroscopicus* or *viridochromogenes*, protein expression vectors normally additionally require an expression cassette which incorporates the cis-acting sequences necessary for transcription and translation of the gene of interest. The cis-acting sequences required for expression in prokaryotes differ from those required in eukaryotes and plants.

A eukaryotic expression cassette requires a transcriptional promoter upstream (5') to the gene of interest, a transcriptional termination region such as a poly-A addition site, and a ribosome binding site upstream of the gene of interest's first codon. In bacterial cells, a useful transcriptional promoter that could be included in the vector is the T7 RNA Polymerase-binding promoter. Promoters, as previously described herein, are known to efficiently promote transcription of mRNA. Also upstream from the gene of interest the vector may include a nucleotide sequence encoding a signal sequence known to direct a covalently linked protein to a particular compartment of the host cells such as the cell surface.

Insect viruses, or baculoviruses, are known to infect and adversely affect certain insects. The affect of the viruses on insects is slow, and viruses do not stop the feeding of insects. Thus viruses are not viewed as being useful as insect pest control agents. Combining the *Photographus* toxins genes into a baculovirus vector could provide an efficient way of transmitting the toxins while increasing the lethality of the virus. In addition, since different baculoviruses are specific to different insects, it may be possible to use a particular toxin to selectively target particularly damaging insect pests. A particularly useful vector for the toxins genes is the nuclear polyhedrosis virus. Transfer vectors using this virus have been described and are now the vectors of choice for transferring foreign genes into insects. The virus-toxin gene recombinant may be constructed in an orally transmissible form. Baculoviruses normally infect insect victims through the mid-gut intestinal mucosa. The toxin gene inserted behind a strong viral coat

protein promoter would be expressed and should rapidly kill the infected insect.

In addition to an insect virus or baculovirus or transgenic plant delivery system for the protein toxins of the present invention, the proteins may be encapsulated using *Bacillus thuringiensis* encapsulation technology such as but not limited to U.S. Patent Nos. 4,695,455; 4,695,462; 4,861,595 which are all incorporated herein by reference. Another delivery system for the protein toxins of the present invention is formulation of the protein into a bait matrix, which could then be used in above and below ground insect bait stations. Examples of such technology include but are not limited to PCT Patent Application WO 93/23998, which is incorporated herein by reference.

As is described above, it might become necessary to modify the sequence encoding the protein when expressing it in a non-native host, since the codon preferences of other hosts may differ from that of *Photobacterium*. In such a case, translation may be quite inefficient in a new host unless compensating modifications to the coding sequence are made. Additionally, modifications to the amino acid sequence might be desirable to avoid inhibitory cross-reactivity with proteins of the new host, or to refine the insecticidal properties of the protein in the new host. A genetically modified toxin gene might encode a toxin exhibiting, for example, enhanced or reduced toxicity, altered insect resistance development, altered stability, or modified target species specificity.

In addition to the *Photobacterium* genes encoding the toxins, the scope of the present invention is intended to include related nucleic acid sequences which encode amino acid biopolymers homologous to the toxin proteins and which retain the toxic effect of the *Photobacterium* proteins in insect species after oral ingestion.

For instance, the toxins used in the present invention seem to first inhibit larval feeding before death ensues. By manipulating the nucleic acid sequence of *Photobacterium* toxins or its controlling sequences, genetic engineers placing the toxin gene into plants could modulate its potency or its mode of action to, for example, keep the eating-inhibitory activity while eliminating the absolute toxicity to the larvae. This change could permit the transformed plant to survive until harvest

without having the unnecessarily dramatic effect on the ecosystem of wiping out all target insects. All such modifications of the gene encoding the toxin, or of the protein encoded by the gene, are envisioned to fall within the scope of the present invention.

5 Other envisioned modifications of the nucleic acid include the addition of targeting sequences to direct the toxin to particular parts of the insect larvae for improving its efficiency.

10 Strains ATCC 55397, 43948, 43949, 43950, 43951, 43952 have been deposited in the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852 USA. Amino acid and nucleotide sequence data for the W-14 native toxin (ATCC 55397) is presented below. Isolation of the genomic DNA for the toxins from the bacterial hosts is also exemplified herein.

15 Standard and molecular biology techniques were followed and taught in the specification herein. Additional information may be found in Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989), Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, which is incorporated herein by reference.

20 The following abbreviations are used throughout the Examples:
Tris = tris (hydroxymethyl) amino methane; SDS = sodium dodecyl sulfate; EDTA = ethylenediaminetetraacetic acid, IPTG = isopropylthio-B-galactoside, X-gal = 5-bromo-4-chloro-3-indoyl-B-D-galactoside, CTAB = cetyltrimethylammonium bromide; kbp =
25 kilobase pairs; dATP, dCTP, dGTP, dTTP, I = 2'-deoxynucleoside 5'-triphosphates of adenine, cytosine, guanine, thymine, and inosine, respectively; ATP = adenosine 5' triphosphate.

30

Example 1

Purification of toxin from *P. luminescens* and Demonstration of toxicity after oral delivery of purified toxin

35 The insecticidal protein toxin of the present invention was purified from *P. luminescens* strain W-14, ATCC Accession Number 55397. Stock cultures of *P. luminescens* were maintained on petri dishes containing 2% Proteose Peptone No. 3 (i.e., PP3, Difco Laboratories, Detroit MI) in 1.5% agar, incubated at 25°C and transferred weekly. Colonies of the primary form of the bacteria
40 were inoculated into 200 ml of PP3 broth supplemented with 0.5%

polyoxyethylene sorbitan mono-stearate (Tween 60, Sigma Chemical Company, St. Louis MO) in a one liter flask. The broth cultures were grown for 72 hours at 30°C on a rotary shaker. The toxin proteins can be recovered from cultures grown in the presence or
5 absence of Tween; however, the absence of Tween can affect the form of the bacteria grown and the profile of proteins produced by the bacteria. In the absence of Tween, a variant shift occurs insofar as the molecular weight of at least one identified toxin subunit shifts from about 200 kDa to about 185 kDa.

10 The 72 hour cultures were centrifuged at 10,000 x g for 30 minutes to remove cells and debris. The supernatant fraction that contained the insecticidal activity was decanted and brought to 50 mM K₂HPO₄ by adding an appropriate volume of 1.0 M K₂HPO₄. The pH was adjusted to 8.6 by adding potassium hydroxide. This
15 supernatant fraction was then mixed with DEAE-Sephacel (Pharmacia LKB Biotechnology) which had been equilibrated with 50 mM K₂HPO₄. The toxic activity was adsorbed to the DEAE resin. This mixture was then poured into a 2.6 x 40 cm column and washed with 50 mM K₂HPO₄ at room temperature at a flow rate of 30 ml/hr until the
20 effluent reached a steady baseline UV absorbance at 280 nm. The column was then washed with 150 mM KCl until the effluent again reached a steady 280 nm baseline. Finally the column was washed with 300 mM KCl and fractions were collected.

Fractions containing the toxin were pooled and filter
25 sterilized using a 0.2 micron pore membrane filter. The toxin was then concentrated and equilibrated to 100 mM KPO₄, pH 6.9, using an ultrafiltration membrane with a molecular weight cutoff of 100 kDa at 4°C (Centriprep 100, Amicon Division-W.R. Grace and Company). A 3 ml sample of the toxin concentrate was applied to
30 the top of a 2.6 x 95 cm Sephacryl S-400 HR gel filtration column (Pharmacia LKB Biotechnology). The eluent buffer was 100 mM KPO₄, pH 6.9, which was run at a flow rate of 17 ml/hr, at 4°C. The effluent was monitored at 280 nm.

Fractions were collected and tested for toxic activity.
35 Toxicity of chromatographic fractions was examined in a biological assay using *Manduca sexta* larvae. Fractions were either applied directly onto the insect diet (Gypsy moth wheat germ diet, ICN Biochemicals Division - ICN Biomedicals, Inc.) or administered by intrahemocelic injection of a 5 ul sample through
40 the first proleg of 4th or 5th instar larva using a 30 gauge

needle. The weight of each larva within a treatment group was recorded at 24 hour intervals. Toxicity was presumed if the insect ceased feeding and died within several days of consuming treated insect diet or if death occurred within 24 hours after
5 injection of a fraction.

The toxic fractions were pooled and concentrated using the Centriprep-100 and were then analyzed by HPLC using a 7.5 mm x 60 cm TSK-GEL G-4000 SW gel permeation column with 100 mM potassium phosphate, pH 6.9 eluent buffer running at 0.4 ml/min. This
10 analysis revealed the toxin protein to be contained within a single sharp peak that eluted from the column with a retention time of approximately 33.6 minutes. This retention time corresponded to an estimated molecular weight of 1,000 kDa. Peak fractions were collected for further purification while fractions
15 not containing this protein were discarded. The peak eluted from the HPLC absorbs UV light at 218 and 280 nm but did not absorb at 405 nm. Absorbance at 405 nm was shown to be an attribute of xenorhabdin antibiotic compounds.

Electrophoresis of the pooled peak fractions in a non-denaturing agarose gel (Metaphor Agarose, FMC BioProducts) showed
20 that two protein complexes are present in the peak. The peak material, buffered in 50 mM Tris-HCl, pH 7.0, was separated on a 1.5% agarose stacking gel buffered with 100 mM Tris-HCl at pH 7.0 and 1.9% agarose resolving gel buffered with 200 mM Tris-borate
25 at pH 8.3 under standard buffer conditions (anode buffer 1M Tris-HCl, pH 8.3; cathode buffer 0.025 M Tris, 0.192 M glycine). The gels were run at 13 mA constant current at 15°C until the phenol red tracking dye reached the end of the gel. Two protein bands were visualized in the agarose gels using Coomassie brilliant
30 blue staining.

The slower migrating band was referred to as "protein band 1" and faster migrating band was referred to as "protein band 2." The two protein bands were present in approximately equal amounts. The Coomassie stained agarose gels were used as a guide
35 to precisely excise the two protein bands from unstained portions of the gels. The excised pieces containing the protein bands were macerated and a small amount of sterile water was added. As a control, a portion of the gel that contained no protein was also excised and treated in the same manner as the gel pieces
40 containing the protein. Protein was recovered from the gel

pieces by electroelution into 100 mM Tris-borate pH 8.3, at 100 volts (constant voltage) for two hours. Alternatively, protein was passively eluted from the gel pieces by adding an equal volume of 50 mM Tris-HCl, pH 7.0, to the gel pieces, then
5 incubating at 30°C for 16 hours. This allowed the protein to diffuse from the gel into the buffer, which was then collected.

Results of insect toxicity tests using HPLC-purified toxin (33.6 min. peak) and agarose gel purified toxin demonstrated toxicity of the extracts. Injection of 1.5 µg of the HPLC
10 purified protein kills within 24 hours. Both protein bands 1 and 2, recovered from agarose gels by passive elution or electroelution, were lethal upon injection. The protein concentration estimated for these samples was less than 50 ng/larva. A comparison of the weight gain and the mortality
15 between the groups of larvae injected with protein bands 1 or 2 indicate that protein band 1 was more toxic by injection delivery.

When HPLC-purified toxin was applied to larval diet at a concentration of 7.5 µg/larva, it caused a halt in larval weight
20 gain (24 larvae tested). The larvae begin to feed, but after consuming only a very small portion of the toxin treated diet they began to show pathological symptoms induced by the toxin and the larvae cease feeding. The insect frass became discolored and most larva showed signs of diarrhea. Significant insect
25 mortality resulted when several 5 µg toxin doses were applied to the diet over a 7-10 day period.

Agarose-separated protein band 1 significantly inhibited larval weight gain at a dose of 200 ng/larva. Larvae fed similar concentrations of protein band 2 were not inhibited and gained
30 weight at the same rate as the control larvae. Twelve larvae were fed eluted protein and 45 larvae were fed protein-containing agarose pieces. These two sets of data indicate that protein band 1 was orally toxic to *Manduca sexta*. In this experiment it appeared that protein band 2 was not toxic to *Manduca sexta*.

35 Further analysis of protein bands 1 and 2 by SDS-PAGE under denaturing conditions showed that each band was composed of several smaller protein subunits. Proteins were visualized by Coomassie brilliant blue staining followed by silver staining to achieve maximum sensitivity.

The protein subunits in the two bands were very similar. Protein band 1 contains 8 protein subunits of 25.1, 56.2, 60.8, 65.6, 166, 171, 184 and 208 kDa. Protein band 2 had an identical profile except that the 25.1, 60.8, and 65.6 kDa proteins were not present. The 56.2, 60.8, 65.6, and 184 kDa proteins were present in the complex of protein band 1 at approximately equal concentrations and represent 80% or more of the total protein content of that complex.

The native HPLC-purified toxin was further characterized as follows. The toxin was heat labile in that after being heated to 60°C for 15 minutes it lost its ability to kill or to inhibit weight gain when injected or fed to *M. sexta* larvae. Assays were designed to detect lipase, type C phospholipase, nuclease or red blood cell hemolysis activities and were performed with purified toxin. None of these activities were present. Antibiotic zone inhibition assays were also done and the purified toxin failed to inhibit growth of Gram-negative or -positive bacteria, yeast or filamentous fungi, indicating that the toxic is not a xenorhabdin antibiotic.

The native HPLC-purified toxin was tested for ability to kill insects other than *Manduca sexta*. Table 2 lists insects killed by the HPLC-purified *P. luminescens* toxin in this study.

Table 2

Insects Killed by *P. luminescens* Toxin

	<u>Common Name</u>	<u>Order</u>	<u>Genus and species</u>	<u>Route of Delivery</u>
30	Tobacco horn worm	Lepidoptera	<i>Manduca sexta</i>	Oral and injected
	Mealworm	Coleoptera	<i>Tenebrio molitor</i>	Oral
35	Pharaoh ant	Hymenoptera	<i>Monomorium pharoanis</i>	Oral
	German cockroach	Dictyoptera	<i>Blattella germanica</i>	Oral and injected
40	Mosquito	Diptera	<i>Aedes aegypti</i>	Oral

Example 2
Insecticide Utility

The *Photobacterium luminescens* utility and toxicity were further characterized. *Photobacterium luminescens* (strain W-14) culture broth was produced as follows. The production medium was 2% Bacto Proteose Peptone[®] Number 3 (PP3, Difco Laboratories, Detroit, Michigan) in Milli-Q[®] deionized water. Seed culture flasks consisted of 175 ml medium placed in a 500 ml tribaffled flask with a Delong neck, covered with a Kaput and autoclaved for 20 minutes, T=250°F. Production flasks consisted of 500 mls in a 2.8 liter 500 ml tribaffled flask with a Delong neck, covered by a Shin-etsu silicon foam closure. These were autoclaved for 45 minutes, T=250°F. The seed culture was incubated at 28°C at 150 rpm in a gyrotory shaking incubator with a 2 inch throw. After 16 hours of growth, 1% of the seed culture was placed in the production flask which was allowed to grow for 24 hours before harvest. Production of the toxin appears to be during log phase growth. The microbial broth was transferred to a 1L centrifuge bottle and the cellular biomass was pelleted (30 minutes at 2500 RPM at 4°C, [R.C.F. = ~1600] HG-4L Rotor RC3 Sorval centrifuge, Dupont, Wilmington, Delaware). The primary broth was chilled at 4°C for 8 - 16 hours and recentrifuged at least 2 hours (conditions above) to further clarify the broth by removal of a putative mucopolysaccharide which precipitated upon standing. (An alternative processing method combined both steps and involved the use of a 16 hour clarification centrifugation, same conditions as above.) This broth was then stored at 4°C prior to bioassay or filtration.

Photobacterium culture broth and protein toxin(s) purified from this broth showed activity (mortality and/or growth inhibition, reduced adult emergence) against a number of insects. More specifically, the activity is seen against corn rootworm (larvae and adult), Colorado potato beetle, and turf grubs, which are members of the insect order Coleoptera. Other members of the Coleoptera include wireworms, pollen beetles, flea beetles, seed beetles and weevils. Activity has also been observed against aster leafhopper, which is a member of the order, Homoptera. Other members of the Homoptera include planthoppers, pear psylla, apple sucker, scale insects, whiteflies, and spittle bugs, as

well as numerous host specific aphid species. The broth and purified fractions are also active against beet armyworm, cabbage looper, black cutworm, tobacco budworm, European corn borer, corn earworm, and codling moth, which are members of the order

5 *Lepidoptera*. Other typical members of this order are clothes moth, Indian mealmoth, leaf rollers, cabbage worm, cotton bollworm, bagworm, Eastern tent caterpillar, sod webworm, and fall armyworm. Activity is also seen against fruitfly and

10 mosquito larvae, which are members of the order *Diptera*. Other members of the order *Diptera* are pea midge, carrot fly, cabbage root fly, turnip root fly, onion fly, crane fly, house fly, and various mosquito species. Activity is seen against carpenter ant and Argentine ant, which are members of the order that also includes fire ants, odorous house ants, and little black ants.

15 The broth/fraction is useful for reducing populations of insects and were used in a method of inhibiting an insect population. The method may comprise applying to a locus of the insect an effective insect inactivating amount of the active described. Results are reported in Table 3.

20 Activity against corn rootworm larvae was tested as follows. *Photorhabdus* culture broth (filter sterilized, cell-free) or purified HPLC fractions were applied directly to the surface (~1.5 cm²) of 0.25 ml of artificial diet in 30 µl aliquots following dilution in control medium or 10 mM sodium phosphate

25 buffer, pH 7.0, respectively. The diet plates were allowed to air-dry in a sterile flow-hood and the wells were infested with single, neonate *Diabrotica undecimpunctata howardi* (Southern corn rootworm, SCR) hatched from sterilized eggs, with second instar SCR grown on artificial diet or with second instar *Diabrotica*

30 *virgifera virgifera* (Western corn rootworm, WCR) reared on corn seedlings grown in Metromix®. Second instar larvae were weighed prior to addition to the diet. The plates were sealed, placed in a humidified growth chamber and maintained at 27°C for the appropriate period (4 days for neonate and adult SCR, 2-5 days

35 for WCR larvae, 7-14 days for second instar SCR). Mortality and weight determinations were scored as indicated. Generally, 16 insects per treatment were used in all studies. Control mortalities were as follows: neonate larvae, <5%, adult beetles, 5%.

Activity against Colorado potato beetle was tested as follows. *Photorhabdus* culture broth or control medium was applied to the surface (~2.0 cm²) of 1.5 ml of standard artificial diet held in the wells of a 24-well tissue culture plate. Each well
5 received 50 µl of treatment and was allowed to air dry.

Individual second instar Colorado potato beetle (*Leptinotarsa decemlineata*, CPB) larvae were then placed onto the diet and mortality was scored after 4 days. Ten larvae per treatment were used in all studies. Control mortality was 3.3%.

10 Activity against Japanese beetle grubs and beetles was tested as follows. Turf grubs (*Popillia japonica*, 2-3rd instar) were collected from infested lawns and maintained in the laboratory in soil/peat mixture with carrot slices added as additional diet. Turf beetles were pheromone-trapped locally and
15 maintained in the laboratory in plastic containers with maple leaves as food. Following application of undiluted *Photorhabdus* culture broth or control medium to corn rootworm artificial diet (30 µl/1.54 cm², beetles) or carrot slices (larvae), both stages were placed singly in a diet well and observed for any mortality
20 and feeding. In both cases there was a clear reduction in the amount of feeding (and feces production) observed.

Activity against mosquito larvae was tested as follows. The assay was conducted in a 96-well microtiter plate. Each well contained 200 µl of aqueous solution (*Photorhabdus* culture broth,
25 control medium or H₂O) and approximately 20, 1-day old larvae (*Aedes aegypti*). There were 6 wells per treatment. The results were read at 2 hours after infestation and did not change over the three day observation period. No control mortality was seen.

Activity against fruitflies was tested as follows.
30 Purchased *Drosophila melanogaster* medium was prepared using 50% dry medium and a 50% liquid of either water, control medium or *Photorhabdus* culture broth. This was accomplished by placing 8.0 ml of dry medium in each of 3 rearing vials per treatment and adding 8.0 ml of the appropriate liquid. Ten late instar
35 *Drosophila melanogaster* maggots were then added to each vial. The vials were held on a laboratory bench, at room temperature, under fluorescent ceiling lights. Pupal or adult counts were made after 3, 7 and 10 days of exposure. Incorporation of *Photorhabdus* culture broth into the diet media for fruitfly

WO 97/17432

PCT/US96/18003

maggots caused a slight (17%) but significant reduction in day-10 adult emergence as compared to water and control medium (3% reduction).

- Activity against aster leafhopper was tested as follows.
- 5 The ingestion assay for aster leafhopper (Macrosteles severini) is designed to allow ingestion of the active "food" solution external contact. The reservoir for the active "food" solution is made by making 2 holes in the center of the bottom portion of a 35 x 10 mm Petri dish. A 2 inch Parafilm M[®] square is placed across the top of the dish and secured with an "O" ring. A 1 oz. plastic cup is then infested with approximately 7 leafhoppers and the reservoir is placed on top of the cup, Parafilm down. The test solution is then added to the reservoir through the holes. In tests using undiluted Photorhabdus culture broth, the broth and control medium were dialyzed against water to reduce control mortality. Mortality is reported at day 2 where 26.5% control mortality was seen. In the tests using purified fractions (200 mg protein/ml) a final concentration of 5% sucrose was used in all treatments to improve survivability of the aster leafhoppers. The assay was graded for mortality at 72 hours.
- 20 Control mortality was 5.5%.
- Activity against Argentine ants was tested as follows. A 1.5 ml aliquot of 100% Photorhabdus culture broth, control medium or water was pipetted into 2.0 ml clear glass vials. The vials were plugged with a piece of cotton dental wick that was moistened with the appropriate treatment. Each vial was placed into a separate 60x16mm Petri dish with 8 to 12 adult Argentine ants (*Linepithema humile*). There were three replicates per treatment. Bioassay plates were held on a laboratory bench, at room temperature under fluorescent ceiling lights. Mortality readings were made after 5 days of exposure. Control mortality was 24%.
- 35 Activity against carpenter ant was tested as follows. Black carpenter ant workers (*Camponotus pennsylvanicus*) were collected from trees on DowElanco property in Indianapolis, IN. Tests with Photorhabdus culture broth were performed as follows. Each plastic bioassay container (7 1/8" x 3") held fifteen workers, a paper harborage and 10 ml of broth or control media in a plastic shot glass. A cotton wick delivered the treatment to the ants

through a hole in the shot glass lid. All treatments contained 5% sucrose. Bioassays were held in the dark at room temperature and graded at 19 days. Control mortality was 9%. Assays delivering purified fractions utilized artificial ant diet mixed with the treatment (purified fraction or control solution) at a rate of 0.2 ml treatment/2.0 g diet in a plastic test tube. The final protein concentration of the purified fraction was less than 10 µg/g diet. Ten ants per treatment, a water source, harborage and the treated diet were placed in sealed plastic containers and maintained in the dark at 27°C in a humidified incubator. Mortality was scored at day 10. No control mortality was seen.

Activity against various lepidopteran larvae was tested as follows. *Photorhabdus* culture broth or purified fractions were applied directly to the surface (~1.5 cm²) of 0.25 ml of standard artificial diet in 30 µl aliquots following dilution in control medium or 10 mM sodium phosphate buffer, pH 7.0, respectively. The diet plates were allowed to air-dry in a sterile flow-hood and the wells were infested with single, neonate larva. European corn borer (*Ostrinia nubilalis*) and corn earworm (*Helicoverpa zea*) eggs were supplied from commercial sources and hatched in-house, whereas beet armyworm (*Spodoptera exigua*), cabbage looper (*Trichoplusia ni*), tobacco budworm (*Heliothis virescens*), codling moth (*Laspeyresia pomonella*) and black cutworm (*Agrotis ipsilon*) larvae were supplied internally. Following infestation with larvae, the diet plates were sealed, placed in a humidified growth chamber and maintained in the dark at 27°C for the appropriate period. Mortality and weight determinations were scored at days 5-7 for *Photorhabdus* culture broth and days 4-7 for the purified fraction. Generally, 16 insects per treatment were used in all studies. Control mortality ranged from 4-12.5% for control medium and was less than 10% for phosphate buffer.

Table 3

Effect of *Photorhabdus luminescens* (strain W-14)
Culture Broth and Purified Toxin Fraction on Mortality and Growth
Inhibition of Different Insect Orders/Species

5

Insect Order/Species	Broth		Purified Fraction	
	% Mort.	% G.I.	% Mort.	% G.I.
COLEOPTERA				
Corn Rootworm				
Southern/neonate larva	100	na	100	na
Southern/2 nd instar	na	38.5	nt	nt
Southern/adult	45	nt	nt	nt
Western/2 nd instar	na	35	nt	nt
Colorado Potato				
Beetle	93	nt	nt	nt
2 nd instar				
Turf Grub	na	a.f.	nt	nt
3 rd instar	na	a.f.	nt	nt
adult				
DIPTERA				
Fruit Fly (adult emergence)	17	nt	nt	nt
	100	na	nt	nt
Mosquito larvae				
HOMOPTERA				
Aster Leafhopper	96.5	na	100	na
HYMENOPTERA				
Argentine Ant	75	na	nt	na
Carpenter Ant	71	na	100	na
LEPIDOPTERA				
Beet Armyworm	12.5	36	18.75	41.4
Black Cutworm	nt	nt	0	71.2
Cabbage Looper	nt	nt	21.9	66.8
Codling Moth	nt	nt	6.25	45.9
Corn Earworm	56.3	94.2	97.9	na
European Corn Borer	96.7	98.4	100	na
Tobacco Budworm	13.5	52.5	19.4	85.6

Mort. = mortality, G.I. = growth inhibition,
na = not applicable, nt = not tested, a.f. = anti-feedant

Example 3Insecticide Utility Upon Soil Application

Photorhabdus luminescens (strain W-14) culture broth was shown to be active against corn rootworm when applied directly to soil or a soil-mix (Metromix®). Activity against neonate SCR and WCR in Metromix® was tested as follows (Table 4). The test was run using corn seedlings (United Agriseeds brand CL614) that were germinated in the light on moist filter paper for 6 days. After roots were approximately 3-6 cm long, a single kernel/seedling was planted in a 591 ml clear plastic cup with 50 gm of dry Metromix®. Twenty neonate SCR or WCR were then placed directly on the roots of the seedling and covered with Metromix®. Upon infestation, the seedlings were then drenched with 50 ml total volume of a diluted broth solution. After drenching, the cups were sealed and left at room temperature in the light for 7 days. Afterwards, the seedlings were washed to remove all Metromix® and the roots were excised and weighed. Activity was rated as the percentage of corn root remaining relative to the control plants and as leaf damage induced by feeding. Leaf damage was scored visually and rated as either -, +, ++, or +++, with - representing no damage and +++ representing severe damage.

Activity against neonate SCR in soil was tested as follows (Table 5). The test was run using corn seedlings (United Agriseeds brand CL614) that were germinated in the light on moist filter paper for 6 days. After the roots were approximately 3-6 cm long, a single kernel/seedling was planted in a 591 ml clear plastic cup with 150 gm of soil from a field in Lebanon, IN planted the previous year with corn. This soil had not been previously treated with insecticides. Twenty neonate SCR were then placed directly on the roots of the seedling and covered with soil. After infestation, the seedlings were drenched with 50 ml total volume of a diluted broth solution. After drenching, the unsealed cups were incubated in a high relative humidity chamber (80%) at 78°F. Afterwards, the seedlings were washed to remove all soil and the roots were excised and weighed. Activity was rated as the percentage of corn root remaining relative to the control plants and as leaf damage induced by feeding. Leaf damage was scored visually and rated as either -, +, ++, or +++, with - representing no damage and +++ representing severe damage.

Table 4
Effect of *Photorhabdus luminescens* (strain W-14) Culture Broth on Rootworm Larvae After Post-Infestation Drenching (Metromix®)

5		Treatment	Larvae	Leaf Damage	Root Weight (g)	%
		Southern Corn Rootworm				
10		Water	-	-	0.4916 ± 0.023	100
		Medium (2.0% v/v)	-	-	0.4416 ± 0.029	100
		Broth (6.25%v/v)	-	-	0.4641 ± 0.081	100
		Water	+	+++	0.1410 ± 0.006	28.7
15		Media (2.0% v/v)	+	+++	0.1345 ± 0.028	30.4
		Broth (1.56% v/v)	+	-	0.4830 ± 0.031	104
		Western Corn Rootworm				
20		Water	-	-	0.4446 ± 0.019	100
		Broth (2.0% v/v)	-	-	0.4069 ± 0.026	100
		Water	+	-	0.2202 ± 0.015	49
25		Broth (2.0% v/v)	+	-	0.3879 ± 0.013	95

Table 5
Effect of *Photorhabdus luminescens* (strain W-14) Culture Broth on Southern Corn Rootworm Larvae After Post-Infestation Drenching (Soil)

30		Treatment	Larvae	Leaf Damage	Root Weight(g)	%
		Water	-	-	0.2148 ± 0.014	100
35		Broth (50% v/v)	-	-	0.2260 ± 0.016	103
		Water	+	+++	0.0916 ± 0.009	43
		Broth (50% v/v)	+	-	0.2428 ± 0.032	113

40 Activity of *Photorhabdus luminescens* (strain W-14) culture broth against second instar turf grubs in Metromix® was observed in tests conducted as follows (Table 6). Approximately 50 gm of dry Metromix® was added to a 591 ml clear plastic cup. The Metromix® was then drenched with 50 ml total volume of a 50% (v/v) 45 diluted *Photorhabdus* broth solution. The dilution of crude broth was made with water, with 50% broth being prepared by adding 25 ml of crude broth to 25 ml of water for 50 ml total volume. A 1% (w/v) solution of proteose peptone #3 (PP3), which is a 50% 50 dilution of the normal media concentration, was used as a broth control. After drenching, five second instar turf grubs were

placed on the top of the moistened Metromix[®]. Healthy turf grub larvae burrowed rapidly into the Metromix[®]. Those larvae that did not burrow within 1h were removed and replaced with fresh larvae. The cups were sealed and placed in a 28°C incubator, in the dark.

- 5 After seven days, larvae were removed from the Metromix[®] and scored for mortality. Activity was rated the percentage of mortality relative to control.

10

Table 6

Effect of *Photorhabdus luminescens* (strain W-14) Culture Broth on Turf Grub After Pre-Infestation Drenching (Metromix[®])

15	Treatment	Mortality*	Mortality %
	Water	7/15	47
	Control medium (1.0% w/v)	12/19	63
20	Broth (50% v/v)	17/20	85

25 *expressed as a ratio of dead/living larvae

Example 4

Insecticide Utility Upon Leaf Application

- 30 Activity of *Photorhabdus* broth against European corn borer was seen when the broth was applied directly to the surface of maize leaves (Table 7). In these assays *Photorhabdus* broth was diluted 100-fold with culture medium and applied manually to the surface of excised maize leaves at a rate of ~6.0 $\mu\text{l}/\text{cm}^2$ of leaf
- 35 surface. The leaves were air dried and cut into equal sized strips approximately 2 x 2 inches. The leaves were rolled, secured with paper clips and placed in 1 oz plastic shot glasses with 0.25 inch of 2% agar on the bottom surface to provide moisture. Twelve neonate European corn borers were then placed
- 40 onto the rolled leaf and the cup was sealed. After incubation for 5 days at 27°C in the dark, the samples were scored for feeding damage and recovered larvae.

Table 7

Effect of *Photorhabdus luminescens* (strain W-14) Culture Broth on European Corn Borer Larvae Following Pre-Infestation Application to Excised Maize Leaves

Treatment	Leaf Damage	Larvae Recovered	Weight (mg)
Water	Extensive	55/120	0.42 mg
Control Medium	Extensive	40/120	0.50 mg
Broth (1.0% v/v)	Trace	3/120	0.15 mg

10

Activity of the culture broth against neonate tobacco budworm (*Heliothis virescens*) was demonstrated using a leaf dip methodology. Fresh cotton leaves were excised from the plant and leaf disks were cut with an 18.5 mm cork-borer. The disks were individually emersed in control medium (PP3) or *Photorhabdus luminescens* (strain W-14) culture broth which had been concentrated approximately 10-fold using an Amicon (Beverly, MA), Proflux M12 tangential filtration system with a 10 kDa filter. Excess liquid was removed and a straightened paper clip was placed through the center of the disk. The paper clip was then wedged into a plastic, 1.0 oz shot glass containing approximately 2.0 ml of 1% Agar. This served to suspend the leaf disk above the agar. Following drying of the leaf disk, a single neonate tobacco budworm larva was placed on the disk and the cup was capped. The cups were then sealed in a plastic bag and placed in a darkened, 27°C incubator for 5 days. At this time the remaining larvae and leaf material were weighed to establish a measure of leaf damage (Table 8).

30

Table 8

Effect of *Photorhabdus luminescens* (Strain W-14) Culture Broth on Tobacco Budworm Neonates in a Cotton-Leaf Dip Assay

Treatment	Leaf Disk	Final Weights (mg) Larvae
Control leaves	55.7 ± 1.3	na*
Control Medium	34.0 ± 2.9	4.3 ± 0.91
<i>Photorhabdus</i> broth	54.3 ± 1.4	0.0**

* - not applicable, ** - no live larvae found

40

Example 5, Part A
Characterization of Toxin Peptide Components

In a subsequent analysis, the toxin protein subunits of the
5 bands isolated as in Example 1 were resolved on a 7% SDS
polyacrylamide electrophoresis gel with a ratio of 30:0.8
(acrylamide:BIS-acrylamide). This gel matrix facilitates better
resolution of the larger proteins. The gel system used to
estimate the Band 1 and Band 2 subunit molecular weights in
10 Example 1 was an 18% gel with a ratio of 38:0.18 (acrylamide:BIS-
acrylamide), which allowed for a broader range of size
separation, but less resolution of higher molecular weight
components.

In this analysis, 10, rather than 8, protein bands were
15 resolved. Table 9 reports the calculated molecular weights of
the 10 resolved bands, and directly compares the molecular
weights estimated under these conditions to those of the prior
example. It is not surprising that additional bands were
detected under the different separation conditions used in this
20 example. Variations between the prior and new estimates of
molecular weight are also to be expected given the differences in
analytical conditions. In the analysis of this example, it is
thought that the higher molecular weight estimates are more
accurate than in Example 1, as a result of improved resolution.
25 However, these are estimates based on SDS PAGE analysis, which
are typically not analytically precise and result in estimates of
peptides and which may have been further altered due to post- and
co-translational modifications.

Amino acid sequences were determined for the N-terminal
30 portions of five of the 10 resolved peptides. Table 9 correlates
the molecular weight of the proteins and the identified
sequences. In SEQ ID NO:2, certain analyses suggest that the
proline at residue 5 may be an asparagine (asn). In SEQ ID NO:3,
certain analyses suggest that the amino acid residues at
35 positions 13 and 14 are both arginine (arg). In SEQ ID NO:4,
certain analyses suggest that the amino acid residue at position
6 may be either alanine (ala) or serine (ser). In SEQ ID NO:5,
certain analyses suggest that the amino acid residue at position
3 may be aspartic acid (asp).

40

Table 9

	EXAMPLE 1 ESTIMATE	NEW ESTIMATE*	SEQ. LISTING
	208	200.2 kDa	SEQ ID NO:1
5	184	175.0 kDa	SEQ ID NO:2
	65.6	68.1 kDa	SEQ ID NO:3
	60.8	65.1 kDa	SEQ ID NO:4
	56.2	58.3 kDa	SEQ ID NO:5
	25.1	23.2 kDa	SEQ ID NO:15
10	*New estimates are based on SDS PAGE and are not based on gene sequences. SDS PAGE is not analytically precise.		

Example 5, Part BCharacterization of Toxin Peptide Components

15

New N-terminal sequence, SEQ ID NO:15, Ala Gln Asp Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr, was obtained by further N-terminal sequencing of peptides isolated from Native HPLC-purified toxin as described in Example 5, Part A, above. This peptide comes from the *tcaA* gene. The peptide labeled TcaA_{ii}, starts at position 254 and goes to position 491, where the TcaA_{iii} peptide starts, SEQ ID NO:4. The estimated size of the peptide based on the gene sequence is 25,240 Da.

25

Example 6Characterization of Toxin Peptide Components

In yet another analysis, the toxin protein complex was re-isolated from the *Photobacterium luminescens* growth medium (after culture without Tween) by performing a 10% - 80% ammonium sulfate precipitation followed by an ion exchange chromatography step (Mono Q) and two molecular sizing chromatography steps. These conditions were like those used in Example 1. During the first molecular sizing step, a second biologically active peak was found at about 100 ± 10 kDa. Based upon protein measurements, this fraction was 20 - 50 fold less active than the larger, or primary, active peak of about 860 ± 100 kDa (native). During this isolation experiment, a smaller active peak of about 325 ± 50 kDa that retained a considerable portion of the starting biological activity was also resolved. It is thought that the 325 kDa peak is related to or derived from the 860 kDa peak.

A 56 kDa protein was resolved in this analysis. The N-terminal sequence of this protein is presented in SEQ ID NO:6. It is noteworthy that this protein shares significant identity and conservation with SEQ ID NO:5 at the N-terminus, suggesting that the two may be encoded by separate members of a gene family and that the proteins produced by each gene are sufficiently similar to both be operable in the insecticidal toxin complex.

A second, prominent 185 kDa protein was consistently present in amounts comparable to that of protein 3 from Table 9, and may be the same protein or protein fragment. The N-terminal sequence of this 185 kDa protein is shown at SEQ ID NO:7.

Additional N-terminal amino acid sequence data were also obtained from isolated proteins. None of the determined N-terminal sequences appear identical to a protein identified in Table 9. Other proteins were present in isolated preparation. One such protein has an estimated molecular weight of 108 kDa and an N-terminal sequence as shown in SEQ ID NO:8. A second such protein has an estimated molecular weight of 80 kDa and an N-terminal sequence as shown in SEQ ID NO:9.

When the protein material in the approximately 325 kDa active peak was analyzed by size, bands of approximately 51, 31, 28, and 22 kDa were observed. As in all cases in which a molecular weight was determined by analysis of electrophoretic mobility, these molecular weights were subject to error effects introduced by buffer ionic strength differences, electrophoresis power differences, and the like. One of ordinary skill would understand that definitive molecular weight values cannot be determined using these standard methods and that each was subject to variation. It was hypothesized that proteins of these sizes are degradation products of the larger protein species (of approximately 200 kDa size) that were observed in the larger primary toxin complex.

Finally, several preparations included a protein having the N-terminal sequence shown in SEQ ID NO:10. This sequence was strongly homologous to known chaperonin proteins, accessory proteins known to function in the assembly of large protein complexes. Although the applicants could not ascribe such an assembly function to the protein identified in SEQ ID NO:10, it was consistent with the existence of the described toxin protein complex that such a chaperonin protein could be involved in its

assembly. Moreover, although such proteins have not directly been suggested to have toxic activity, this protein may be important to determining the overall structural nature of the protein toxin, and thus, may contribute to the toxic activity or durability of the complex *in vivo* after oral delivery.

Subsequent analysis of the stability of the protein toxin complex to proteinase K was undertaken. It was determined that after 24 hour incubation of the complex in the presence of a 10-fold molar excess of proteinase K, activity was virtually eliminated (mortality on oral application dropped to about 5%). These data confirm the proteinaceous nature of the toxin.

The toxic activity was also retained by a dialysis membrane, again confirming the large size of the native toxin complex.

15

Example 7

Isolation, Characterization and Partial Amino Acid Sequencing of *Photorhabdus* Toxins

Isolation and N-Terminal Amino Acid Sequencing: In a set of experiments conducted in parallel to Examples 5 and 6, ammonium sulfate precipitation of *Photorhabdus* proteins was performed by adjusting *Photorhabdus* broth, typically 2-3 liters, to a final concentration of either 10% or 20% by the slow addition of ammonium sulfate crystals. After stirring for 1 hour at 4°C, the material was centrifuged at 12,000 x g for 30 minutes. The supernatant was adjusted to 80% ammonium sulfate, stirred at 4°C for 1 hour, and centrifuged at 12,000 x g for 60 minutes. The pellet was resuspended in one-tenth the volume of 10 mM Na₂·PO₄, pH 7.0 and dialyzed against the same phosphate buffer overnight at 4°C. The dialyzed material was centrifuged at 12,000 x g for 1 hour prior to ion exchange chromatography.

A HR 16/50 Q Sepharose (Pharmacia) anion exchange column was equilibrated with 10 mM Na₂·PO₄, pH 7.0. Centrifuged, dialyzed ammonium sulfate pellet was applied to the Q Sepharose column at a rate of 1.5 ml/min and washed extensively at 3.0 ml/min with equilibration buffer until the optical density (O.D. 280) reached less than 0.100. Next, either a 60 minute NaCl gradient ranging from 0 to 0.5 M at 3 ml/min, or a series of step elutions using 0.1 M, 0.4 M and finally 1.0 NaCl for 60 minutes each was applied to the column. Fractions were pooled and concentrated using a

Centriprep 100. Alternatively, proteins could be eluted by a single 0.4 M NaCl wash without prior elution with 0.1 M NaCl.

Two milliliter aliquots of concentrated Q Sepharose samples were loaded at 0.5 ml/min onto a HR 16/50 Superose 12 (Pharmacia) gel filtration column equilibrated with 10 mM Na_2PO_4 , pH 7.0. The column was washed with the same buffer for 240 min at 0.5 ml/min and 2 min samples were collected. The void volume material was collected and concentrated using a Centriprep 100. Two milliliter aliquots of concentrated Superose 12 samples were loaded at 0.5 ml/min onto a HR 16/50 Sepharose 4B-CL (Pharmacia) gel filtration column equilibrated with 10 mM Na_2PO_4 , pH 7.0. The column was washed with the same buffer for 240 min at 0.5 ml/min and 2 min samples were collected.

The excluded protein peak was subjected to a second fractionation by application to a gel filtration column that used a Sepharose CL-4B resin, which separates proteins ranging from ~30 kDa to 1000 kDa. This fraction was resolved into two peaks; a minor peak at the void volume (>1000 kDa) and a major peak which eluted at an apparent molecular weight of about 860 kDa. Over a one week period subsequent samples subjected to gel filtration showed the gradual appearance of a third peak (approximately 325 kDa) that seemed to arise from the major peak, perhaps by limited proteolysis. Bioassays performed on the three peaks showed that the void peak had no activity, while the 860 kDa toxin complex fraction was highly active, and the 325 kDa peak was less active, although quite potent. SDS PAGE analysis of Sepharose CL-4B toxin complex peaks from different fermentation productions revealed two distinct peptide patterns, denoted "P" and "S". The two patterns had marked differences in the molecular weights and concentrations of peptide components in their fractions. The "S" pattern, produced most frequently, had 4 high molecular weight peptides (> 150 kDa) while the "P" pattern had 3 high molecular weight peptides. In addition, the "S" peptide fraction was found to have 2-3 fold more activity against European Corn Borer. This shift may be related to variations in protein expression due to age of inoculum and/or other factors based on growth parameters of aged cultures.

Milligram quantities of peak toxin complex fractions determined to be "P" or "S" peptide patterns were subjected to preparative SDS PAGE, and transblotted with TRIS-glycine

(Seprabuff™ to PVDF membranes (ProBlott™, Applied Biosystems) for 3-4 hours. Blots were sent for amino acid analysis and N-terminal amino acid sequencing at Harvard MicroChem and Cambridge ProChem, respectively. Three peptides in the "S" pattern had
5 unique N-terminal amino acid sequences compared to the sequences identified in the previous example. A 201 kDa (TcdA_{ii}) peptide set forth as SEQ ID NO:13 below shared between 33% amino acid identity and 50% similarity with SEQ ID NO:1 (TcbA_{ii}) (Table 10, in Table 10 vertical lines denote amino acid identities and
10 colons indicate conservative amino acid substitutions). A second peptide of 197 kDa, SEQ ID NO:14 (TcdB), had 42% identity and 58% homology with SEQ ID NO:2 (TcaC). Yet a third peptide of 205 kDa was denoted TcdA_{ii}. In addition, a limited N-terminal amino acid sequence, SEQ ID NO:16 (TcbA), of a peptide of at least 235 kDa
15 was identical in homology with the amino acid sequence, SEQ ID NO:12, deduced from a cloned gene (*tcbA*), SEQ ID NO:11, containing a deduced amino acid sequence corresponding to SEQ ID NO:1 (TcbA_{ii}). This indicates that the larger 235+ kDa peptide was proteolytically processed to the 201 kDa peptide, (TcbA_{ii}),
20 (SEQ ID NO:1) during fermentation, possibly resulting in activation of the molecule. In yet another sequence, the sequence originally reported as SEQ ID NO:5 (TcaB_{ii}) reported in Example 5 above, was found to contain an aspartic acid residue (Asp) at the third position rather than glycine (Gly) and two
25 additional amino acids Gly and Asp at the eighth and ninth positions, respectively. In yet two other sequences, SEQ ID NO:2 (TcaC) and SEQ ID NO:3 (TcaB_i), additional amino acid sequence was obtained. Densitometric quantitation was performed using a sample that was identical to the "S" preparation sent for N-
30 terminal analysis. This analysis showed that the 201 kDa and 197 kDa peptides represent 7.0% and 7.2%, respectively, of the total Coomassie brilliant blue stained protein in the "S" pattern and are present in amounts similar to the other abundant peptides. It is speculated that these peptides may represent protein
35 homologs, analogous to the situation found with other bacterial toxins, such as various CryI Bt toxins. These proteins vary from 40-90% homology at their N-terminal amino acid sequence, which encompasses the toxic fragment.

Internal Amino Acid Sequencing: To facilitate cloning of toxin peptide genes, internal amino acid sequences of selected peptides were obtained as followed. Milligram quantities of peak 2A fractions determined to be "P" or "S" peptide patterns were subjected to preparative SDS PAGE, and transblotted with TRIS-glycine (Seprabuff™ to PVDF membranes (ProBlott™, Applied Biosystems) for 3-4 hours. Blots were sent for amino acid analysis and N-terminal amino acid sequencing at Harvard MicroChem and Cambridge ProChem, respectively. Three peptides, referred to as TcbAii (containing SEQ ID NO:1), TcdAii, and TcaB_i (containing SEQ ID NO:3) were subjected to trypsin digestion by Harvard MicroChem followed by HPLC chromatography to separate individual peptides. N-terminal amino acid analysis was performed on selected tryptic peptide fragments. Two internal peptides were sequenced for the peptide TcaB_i (205 kDa peptide) referred to as TcaB_i-PT111 (SEQ ID NO:17) and TcaB_i-PT79 (SEQ ID NO:18). Two internal peptides were sequenced for the peptide TcaB_i (68 kDa peptide) referred to as TcaB_i-PT158 (SEQ ID NO:19) and TcaB_i-PT108 (SEQ ID NO:20). Four internal peptides were sequenced for the peptide TcbAii (201 kDa peptide) referred to as TCBAII-PT103 (SEQ ID NO:21), TcbAii-PT56 (SEQ ID NO:22), TcbAii-PT81(a) (SEQ ID NO:23), and TcbAii-PT81(b) (SEQ ID NO:24).

Table 10

N-Terminal Amino Acid Sequences

	201 kDa (33% identity & 50% similarity to SEQ ID NO.1)	
	L I G Y N N Q F S G * A	SEQ ID NO:13
	: :	
30	F I Q G Y S D L F G N - A	SEQ ID NO:1
	197 kDa (42% identity & 58% similarity SEQ ID NO.2)	
	M Q N S Q T F S V G E L	SEQ ID NO.14
	: : :	
35	M Q D S P E V S I T T L	SEQ ID NO.2

Example 8

40 Construction of a cosmid library of *Photobacterium luminescens* W-14 genomic DNA and its screening to isolate genes encoding peptides comprising the toxic protein preparation

As a prerequisite for the production of *Photobacterium* insect toxic proteins in heterologous hosts, and for other uses, it is necessary to isolate and characterize the genes that encode those

peptides. This objective was pursued in parallel. One approach, described later, was based on the use of monoclonal and polyclonal antibodies raised against the purified toxin which were then used to isolate clones from an expression library. The other approach, described in this example, is based on the use of the N-terminal and internal amino acid sequence data to design degenerate oligonucleotides for use in PCR amplification. Either method can be used to identify DNA clones that contain the peptide-encoding genes so as to permit the isolation of the respective genes, and the determination of their DNA base sequence.

GENOMIC DNA ISOLATION: *Photorhabdus luminescens* strain W-14 (ATCC accession number 55397) was grown on 2% proteose peptone #3 agar (Difco Laboratories, Detroit, MI) and insecticidal toxin competence was maintained by repeated bioassay after passage, using the method described in Example 1 above. A 50 ml shake culture was produced in a 175 ml baffled flask in 2% proteose peptone #3 medium, grown at 28°C and 150 rpm for approximately 24 hours. 15 ml of this culture was pelleted and frozen in its medium at -20°C until it was thawed for DNA isolation. The thawed culture was centrifuged, (700 x g, 30 min) and the floating orange mucopolysaccharide material was removed. The remaining cell material was centrifuged (25,000 x g, 15 min) to pellet the bacterial cells, and the medium was removed and discarded.

Genomic DNA was isolated by an adaptation of the CTAB method described in section 2.4.1 of Current Protocols in Molecular Biology (Ausubel et al. eds, John Wiley & Sons, 1994) [modified to include a salt shock and with all volumes increased 10-fold]. The pelleted bacterial cells were resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to a final volume of 10 ml, then 12 ml of 5 M NaCl was added; this mixture was centrifuged 20 min at 15,000 x g. The pellet was resuspended in 5.7 ml TE and 300 ml of 10% SDS and 60 ml of 20 mg/ml proteinase K (Gibco BRL Products, Grand Island, NY; in sterile distilled water) were added to the suspension. This mixture was incubated at 37°C for 1 hr; then approximately 10 mg lysozyme (Worthington Biochemical Corp., Freehold, NJ) was added. After an additional 45 min, 1 ml of 5 M NaCl and 800 ml of CTAB/NaCl solution (10% w/v CTAB, 0.7 M

NaCl) were added. This preparation was incubated 10 min at 65°C, then gently agitated and further incubated and agitated for approximately 20 min to assist clearing of the cellular material. An equal volume of chloroform/isoamyl alcohol solution (24:1, v/v) was added, mixed gently and centrifuged. After two extractions with an equal volume of PCI (phenol/chloroform/isoamyl alcohol; 50:49:1, v/v/v; equilibrated with 1 M Tris-HCl, pH 8.0; Intermountain Scientific Corporation, Kaysville, UT), the DNA was precipitated with 0.6 volume of isopropanol. The DNA precipitate was gently removed with a glass rod, washed twice with 70% ethanol, dried, and dissolved in 2 ml STE (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 mM EDTA). This preparation contained 2.5 mg/ml DNA, as determined by optical density at 260 nm (i.e., OD₂₆₀).

The molecular size range of the isolated genomic DNA was evaluated for suitability for library construction. CHEF gel analysis was performed in 1.5% agarose (Seakem® LE, FMC BioProducts, Rockland, ME) gels with 0.5 X TBE buffer (44.5 mM Tris-HCl pH 8.0, 44.5 mM H₃BO₃, 1 mM EDTA) on a BioRad CHEF-DR II apparatus with a Pulsewave 760 Switcher (Bio-Rad Laboratories, Inc., Richmond, CA). The running parameters were: initial A time, 3 sec; final A time, 12 sec; 200 volts; running temperature, 4-18°C; run time, 16.5 hr. Ethidium bromide staining and examination of the gel under ultraviolet light indicated the DNA ranged from 30-250 kbp in size.

CONSTRUCTION OF LIBRARY: A partial Sau3A I digest was made of this *Photorhabdus* genomic DNA preparation. The method was based on section 3.1.3 of Ausubel (supra.). Adaptions included running smaller scale reactions under various conditions until nearly optimal results were achieved. Several scaled-up large reactions with varied conditions were run, the results analyzed on CHEF gels, and only the best large scale preparation was carried forward. In the optimal case, 200 µg of *Photorhabdus* genomic DNA was incubated with 1.5 units of Sau3A I (New England Biolabs, "NEB", Beverly, MA) for 15 min at 37°C in 2 ml total volume of 1X NEB 4 buffer (supplied as 10X by the manufacturer). The reaction was stopped by adding 2 ml of PCI and centrifuging at 8000 x g for 10 min. To the supernatant were added 200 µl of 5 M NaCl plus 6 ml of ice-cold ethanol. This preparation was

chilled for 30 min at -20 C, then centrifuged at 12,000 x g for 15 min. The supernatant was removed and the precipitate was dried in a vacuum oven at 40°C, then resuspended in 400 µl STE. Spectrophotometric assay indicated about 40% recovery of the input DNA. The digested DNA was size fractionated on a sucrose gradient according to section 5.3.2 of CPMB (op. cit.). A 10% to 40% (w/v) linear sucrose gradient was prepared with a gradient maker in Ultra-Clear™ tubes (Beckman Instruments, Inc., Palo Alto, CA) and the DNA sample was layered on top. After centrifugation, (26,000 rpm, 17 hr, Beckman SW41 rotor, 20°C), fractions (about 750 µl) were drawn from the top of the gradient and analyzed by CHEF gel electrophoresis (as described earlier). Fractions containing Sau3A 1 fragments in the size range 20-40 kbp were selected and DNA was precipitated by a modification (amounts of all solutions increased approximately 6.3-fold) of the method in section 5.3.3 of Ausubel (supra.). After overnight precipitation, the DNA was collected by centrifugation (17,000 x g, 15 min), dried, redissolved in TE, pooled into a final volume of 80 µl, and reprecipitated with the addition of 8 µl 3 M sodium acetate and 220 µl ethanol. The pellet collected by centrifugation as above was resuspended in 12 µl TE. Concentration of the DNA was determined by Hoechst 33258 dye (Polysciences, Inc., Warrington, PA) fluorometry in a Hoefer TK0100 fluorimeter (Hoefer Scientific Instruments, San Francisco, CA). Approximately 2.5 µg of the size-fractionated DNA was recovered.

Thirty µg of cosmid pWE15 DNA (Stratagene, La Jolla, CA) was digested to completion with 100 units of restriction enzyme BamH 1 (NEB) in the manufacturer's buffer (final volume of 200 µl, 37°C, 1 hr). The reaction was extracted with 100 µl of PCI and DNA was precipitated from the aqueous phase by addition of 20 µl 3M sodium acetate and 550 µl -20°C absolute ethanol. After 30 min at -70°C, the DNA was collected by centrifugation (17,000 x g, 15 min), dried under vacuum, and dissolved in 180 µl of 10 mM Tris-HCl, pH 8.0. To this were added 20 µl of 10X CIP buffer (100 mM Tris-HCl, pH 8.3; 10 mM ZnCl₂; 10 mM MgCl₂), and 1 µl (0.25 units) of 1:4 diluted calf intestinal alkaline phosphatase

(Boehringer Mannheim Corporation, Indianapolis, IN). After 30 min at 37°C, the following additions were made: 2 µl 0.5 M EDTA, pH 8.0; 10 µl 10% SDS; 0.5 µl of 20 mg/ml proteinase K (as above), followed by incubation at 55°C for 30 min. Following sequential extractions with 100 µl of PCI and 100 µl phenol (Intermountain Scientific Corporation, equilibrated with 1 M Tris-HCl, pH 8.0), the dephosphorylated DNA was precipitated by addition of 72 µl of 7.5 M ammonium acetate and 550 µl -20°C ethanol, incubation on ice for 30 min, and centrifugation as above. The pelleted DNA was washed once with 500 µl -20°C 70% ethanol, dried under vacuum, and dissolved in 20 µl of TE buffer.

Ligation of the size-fractionated Sau3A 1 fragments to the BamH 1-digested and phosphatased pWE15 vector was accomplished using T4 ligase (NEB) by a modification (i.e., use of premixed 10X ligation buffer supplied by the manufacturer) of the protocol in section 3.33 of Ausubel. Ligation was carried out overnight in a total volume of 20 µl at 15°C, followed by storage at -20°C.

Four µl of the cosmid DNA ligation reaction, containing about 1 µg of DNA, was packaged into bacteriophage lambda using a commercial packaging extract (Gigapack[®] III Gold Packaging Extract, Stratagene), following the manufacturer's directions. The packaged preparation was stored at 4°C until use. The packaged cosmid preparation was used to infect *Escherichia coli* XL1 Blue MR cells (Stratagene) according to the Gigapack[®] III Gold protocols ("Titering the Cosmid Library"), as follows. XL1 Blue MR cells were grown in LB medium (g/L: Bacto-tryptone, 10; Bacto-yeast extract, 5; Bacto-agar, 15; NaCl, 5; [Difco Laboratories, Detroit, MI]) containing 0.2% (w/v) maltose plus 10 mM MgSO₄, at 37°C. After 5 hr growth, cells were pelleted at 700 x g (15 min) and resuspended in 6 ml of 10 mM MgSO₄. The culture density was adjusted with 10 mM MgSO₄ to OD₆₀₀ = 0.5. The packaged cosmid library was diluted 1:10 or 1:20 with sterile SM medium (0.1 M NaCl, 10 mM MgSO₄, 50 mM Tris-HCl pH 7.5, 0.01% w/v gelatin), and 25 µl of the diluted preparation was mixed with 25 µl of the diluted XL1 Blue MR cells. The mixture was incubated at 25°C for 30 min (without shaking), then 200 µl of LB broth was added, and incubation was continued for approximately 1 hr with occasional

gentle shaking. Aliquots (20-40 μ l) of this culture were spread on LB agar plates containing 100 mg/l ampicillin (i.e., LB-Amp₁₀₀) and incubated overnight at 37°C. To store the library without amplification, single colonies were picked and inoculated into individual wells of sterile 96-well microwell plates; each well containing 75 μ l of Terrific Broth (TB media: 12 g/l Bacto-tryptone, 24 g/l Bacto-yeast extract, 0.4% v/v glycerol, 17 mM KH₂PO₄, 72 mM K₂HPO₄) plus 100 mg/l ampicillin (i.e., TB-Amp₁₀₀) and incubated (without shaking) overnight at 37°C. After replicating the 96-well plate into a copy plate, 75 μ l/well of filter-sterilized TB:glycerol (1:1, v/v; with, or without, 100 mg/l ampicillin) was added to the plate, it was shaken briefly at 100 rpm, 37°C, and then closed with Parafilm® (American National Can, Greenwich, CT) and placed in a -70°C freezer for storage. Copy plates were grown and processed identically to the master plates. A total of 40 such master plates (and their copies) were prepared.

SCREENING OF THE LIBRARY WITH RADIOLABELED DNA PROBES: To prepare colony filters for probing with radioactively labeled probes, ten 96-well plates of the library were thawed at 25°C (bench top at room temperature). A replica plating tool with 96 prongs was used to inoculate a fresh 96-well copy plate containing 75 μ l/well of TB-Amp₁₀₀. The copy plate was grown overnight (stationary) at 37°C, then shaken about 30 min at 100 rpm at 37°C. A total of 800 colonies was represented in these copy plates, due to nongrowth of some isolates. The replica tool was used to inoculate duplicate impressions of the 96-well arrays onto Magna NT (MSI, Westboro, MA) nylon membranes (0.45 micron, 220 x 250 mm) which had been placed on solid LB-Amp₁₀₀ (100 ml/dish) in Bio-assay plastic dishes (Nunc, 243 x 243 x 18 mm; Curtin Mathison Scientific, Inc., Wood Dale, IL). The colonies were grown on the membranes at 37°C for about 3 hr.

A positive control colony (a bacterial clone containing a GZ4 sequence insert, see below) was grown on a separate Magna NT membrane (Nunc, 0.45 micron, 82 mm circle) on LB medium supplemented with 35 mg/l chloramphenicol (i.e., LB-Cam₁₅), and processed alongside the library colony membranes. Bacterial colonies on the membranes were lysed, and the DNA was denatured

and neutralized according to a protocol taken from the Genius™ System User's Guide version 2.0 (Boehringer Mannheim, Indianapolis, IN). Membranes were placed colony side up on filter paper soaked with 0.5 N NaOH plus 1.5 M NaCl for 15 min to denature, and neutralized on filter paper soaked with 1 M Tris-HCl pH 8.0, 1.5 M NaCl for 15 min. After UV-crosslinking using a Stratagene UV Stratalinker set on auto crosslink, the membranes were stored dry at 25°C until use. Membranes were trimmed into strips containing the duplicate impressions of a single 96-well plate, then washed extensively by the method of section 6.4.1 in CPMB (*op. cit.*): 3 hr at 25°C in 3X SSC, 0.1% (w/v) SDS, followed by 1 hr at 65°C in the same solution, then rinsed in 2X SSC in preparation for the hybridization step (20X SSC = 3 M NaCl, 0.3 M sodium citrate, pH 7.0).

15

Amplification of a specific genomic fragment of a tcaC gene.

Based on the N-terminal amino acid sequence determined for the purified TcaC peptide fraction [disclosed herein as SEQ ID NO:2], a pool of degenerate oligonucleotides (pool S4Psh) was synthesized by standard β -cyanoethyl chemistry on an Applied BioSystem ABI394 DNA/RNA Synthesizer (Perkin Elmer, Foster City, CA). The oligonucleotides were deprotected 8 hours at 55°C, dissolved in water, quantitated by spectrophotometric measurement, and diluted for use. This pool corresponds to the determined N-terminal amino acid sequence of the TcaC peptide. The determined amino acid sequence and the corresponding degenerate DNA sequence are given below, where A, C, G, and T are the standard DNA bases, and I represents inosine:

Amino	Met	Gln	Asp	Ser	Pro	Glu	Val
Acid							

S4Psh 5' ATG CA(A/G) GA(T/C) (T/A)(C/G)(T/A) CCI GA(A/G) GT 3'

Another set of degenerate oligonucleotides was synthesized (pool P2.3.5R), representing the complement of the coding strand for the determined amino acid sequence of the SEQ ID NO:17:

Amino	Ala	Phe	Asn	Ile	Asp	Asp	Val
Acid							

Codons 5' GCN TT(T/C) AA(T/C) AT(A/T/C) GA(T/C) GA(T/C) GT 3'
P2.3.5R 3'CG(A/C/G/T) AA(A/G) TT(A/G) TA(T/A/G) CT(A/G) CT(A/G) CA 5'

These oligonucleotides were used as primers in Polymerase Chain Reactions (PCR®, Roche Molecular Systems, Branchburg, NJ) to

amplify a specific DNA fragment from genomic DNA prepared from *Photorhabdus* strain W-14 (see above). A typical reaction (50 μ l) contained 125 pmol of each primer pool P2Psh and P2.3.5R, 253 ng of genomic template DNA, 10 nmol each of dATP, dCTP, dGTP, and dTTP, 1X GeneAmp[®] PCR buffer, and 2.5 units of AmpliTaq[®] DNA polymerase (both from Roche Molecular Systems; 10X GeneAmp[®] buffer is 100 mM Tris-HCl pH 8.3, 500 mM KCl, 0.01% w/v gelatin). Amplifications were performed in a Perkin Elmer Cetus DNA Thermal Cycler (Perkin Elmer, Foster City, CA) using 35 cycles of 94°C (1.0 min), 55°C (2.0 min), 72°C (3.0 min), followed by an extension period of 7.0 min at 72°C. Amplification products were analyzed by electrophoresis through 2% w/v NuSieve[®] 3:1 agarose (FMC BioProducts) in TEA buffer (40 mM Tris-acetate, 2 mM EDTA, pH 8.0). A specific product of estimated size 250 bp was observed amongst numerous other amplification products by ethidium bromide (0.5 μ g/ml) staining of the gel and examination under ultraviolet light.

The region of the gel containing an approximately 250 bp product was excised, and a small plug (0.5 mm dia.) was removed and used to supply template for PCR amplification (40 cycles). The reaction (50 μ l) contained the same components as above, minus genomic template DNA. Following amplification, the ends of the fragments were made blunt and were phosphorylated by incubation at 25°C for 20 min with 1 unit of T4 DNA polymerase (NEB), 1 nmol ATP, and 2.15 units of T4 kinase (Pharmacia Biotech Inc., Piscataway, NJ).

DNA fragments were separated from residual primers by electrophoresis through 1% w/v GTG[®] agarose (FMC) in TEA. A gel slice containing fragments of apparent size 250 bp was excised, and the DNA was extracted using a Qiaex kit (Qiagen Inc., Chatsworth, CA).

The extracted DNA fragments were ligated to plasmid vector pBC KS(+) (Stratagene) that had been digested to completion with restriction enzyme Sma I and extracted in a manner similar to that described for pWE15 DNA above. A typical ligation reaction (16.3 μ l) contained 100 ng of digested pBC KS(+) DNA, 70 ng of 250 bp fragment DNA, 1 nmol [Co(NH₃)₆]Cl₂, and 3.9 Weiss units of T4 DNA ligase (Collaborative Biomedical Products, Bedford, MA), in 1X ligation buffer (50 mM Tris-HCl, pH 7.4; 10 mM MgCl₂; 10 mM

- dithiothreitol; 1 mM spermidine, 1 mM ATP, 100 mg/ml bovine serum albumin). Following overnight incubation at 14°C, the ligated products were transformed into frozen, competent *Escherichia coli* DH5α cells (Gibco BRL) according to the suppliers' recommendations, and plated on LB-Camp plates, containing IPTG (119 µg/ml) and X-gal (50 µg/ml). Independent white colonies were picked, and plasmid DNA was prepared by a modified alkaline-lysis/PEG precipitation method (PRISM™ Ready Reaction DyeDeoxy™ Terminator Cycle Sequencing Kit Protocols; ABI/Perkin Elmer).
- 10 The nucleotide sequence of both strands of the insert DNA was determined, using T7 primers [pBC KS(+) bases 601-623: TAAAACGACGGCCAGTGAAGCGCG] and LacZ primers [pBC KS(+) bases 792-816: ATGACCATGATTACGCCAAGCGCGC] and protocols supplied with the PRISM™ sequencing kit (ABI/Perkin Elmer). Nonincorporated dye-
- 15 terminator dideoxyribonucleotides were removed by passage through Centri-Sep 100 columns (Princeton Separations, Inc., Adelphia, NJ) according to the manufacturer's instructions. The DNA sequence was obtained by analysis of the samples on an ABI Model 373A DNA Sequencer (ABI/Perkin Elmer). The DNA sequences of two
- 20 isolates, GZ4 and HB14, were found to be as illustrated in Figure 1.

- This sequence illustrates the following features: i) bases 1-20 represent one of the 64 possible sequences of the S4Psh degenerate oligonucleotides, ii) the sequence of amino acids 1-3
- 25 and 6-12 correspond exactly to that determined for the N-terminus of TcaC (disclosed as SEQ ID NO:2), iii) the fourth amino acid encoded is a cysteine residue rather than serine. This difference is encoded within the degeneracy for the serine codons (see above), iv) the fifth amino acid encoded is proline,
- 30 corresponding to the TcaC N-terminal sequence given as SEQ ID NO:2, v) bases 257-276 encode one of the 192 possible sequences designed into the degenerate pool, vi) the TGA termination codon introduced at bases 268-270 is the result of complementarity to the degeneracy built into the oligonucleotide pool at the
- 35 corresponding position, and does not indicate a shortened reading frame for the corresponding gene.

Labeling of a TcaC peptide gene-specific probe. DNA fragments corresponding to the above 276 bases were amplified (35

cycles) by PCR[®] in a 100 µl reaction volume, using 100 pmol each of P2Psh and P2.3.5R primers, 10 ng of plasmids GZ4 or HB14 as templates, 20 nmol each of dATP, dCTP, dGTP, and dTTP, 5 units of AmpliTaq[®] DNA polymerase, and 1X concentration of GeneAmp[®] buffer, under the same temperature regimes as described above. The amplification products were extracted from a 1% GTG[®] agarose gel by Qiaex kit and quantitated by fluorometry.

The extracted amplification products from plasmid HB14 template (approximately 400 ng) were split into five aliquots and labeled with ³²P-dCTP using the High Prime Labeling Mix (Boehringer Mannheim) according to the manufacturer's instructions. Nonincorporated radioisotope was removed by passage through NucTrap[®] Probe Purification Columns (Stratagene), according to the supplier's instructions. The specific activity of the labeled DNA product was determined by scintillation counting to be 3.11×10^8 dpm/µg. This labeled DNA was used to probe membranes prepared from 800 members of the genomic library.

Screening with a TcaC-peptide gene specific probe. The radiolabeled HB14 probe was boiled approximately 10 min, then added to "minimal hyb" solution. [Note: The "minimal hyb" method is taken from a CERES protocol; "Restriction Fragment Length Polymorphism Laboratory Manual version 4.0", sections 4-40 and 4-47; CERES/NPI, Salt Lake City, UT. NPI is now defunct, with its successors operating as Linkage Genetics]. "Minimal hyb" solution contains 10% w/v PEG (polyethylene glycol, M.W. approx. 8000), 7% w/v SDS, 0.6X SSC, 10 mM sodium phosphate buffer (from a 1M stock containing 95 g/l NaH₂PO₄·1H₂O and 84.5 g/l Na₂HPO₄·7H₂O), 5 mM EDTA, and 100 mg/ml denatured salmon sperm DNA. Membranes were blotted dry briefly then, without prehybridization, 5 strips of membrane were placed in each of 2 plastic boxes containing 75 ml of "minimal hyb" and 2.6 ng/ml of radiolabeled HB14 probe. These were incubated overnight with slow shaking (50 rpm) at 60°C. The filters were washed three times for approximately 10 min each at 25°C in "minimal hyb wash solution" (0.25X SSC, 0.2% SDS), followed by two 30-min washes with slow shaking at 60°C in the same solution. The filters were placed on paper covered with Saran Wrap[®] (Dow Brands, Indianapolis, IN) in a light-tight autoradiographic cassette and exposed to X-Omat X-ray film (Kodak, Rochester, NY) with two

DuPont Cronex Lightning-Plus C1 enhancers (Sigma Chemical Co., St. Louis, MO), for 4 hr at -70°C. Upon development (standard photographic procedures), significant signals were evident in both replicates amongst a high background of weaker, more irregular signals. The filters were again washed for about 4 hr at 68°C in "minimal hyb wash solution" and then placed again in the cassettes and film was exposed overnight at -70°C. Twelve possible positives were identified due to strong signals on both of the duplicate 96-well colony impressions. No signal was seen with negative control membranes (colonies of XL1 Blue MR cells containing pWE15), and a very strong signal was seen with positive control membranes (DH5α cells containing the GZ4 isolate of the PCR product) that had been processed concurrently with the experimental samples.

The twelve putative hybridization-positive colonies were retrieved from the frozen 96-well library plates and grown overnight at 37°C on solid LB-Amp₁₀₀ medium. They were then patched (3/plate, plus three negative controls: XL1 Blue MR cells containing the pWE15 vector) onto solid LB-Amp₁₀₀. Two sets of membranes (Magna NT nylon, 0.45 micron) were prepared for hybridization. The first set was prepared by placing a filter directly onto the colonies on a patch plate, then removing it with adherent bacterial cells, and processing as below. Filters of the second set were placed on plates containing LB-Amp₁₀₀ medium, then inoculated by transferring cells from the patch plates onto the filters. After overnight growth at 37°C, the filters were removed from the plates and processed.

Bacterial cells on the filters were lysed and DNA denatured by placing each filter colony-side-up on a pool (1.0 ml) of 0.5 M NaOH in a plastic plate for 3 min. The filters were blotted dry on a paper towel, then the process was repeated with fresh 0.5 M NaOH. After blotting dry, the filters were neutralized by placing each on a 1.0 ml pool of 1 M Tris-HCl, pH 7.5 for 3 min, blotted dry, and reneutralised with fresh buffer. This was followed by two similar soakings (5 min each) on pools of 0.5 M Tris-HCl pH 7.5 plus 1.5 M NaCl. After blotting dry, the DNA was UV crosslinked to the filter (as above), and the filters were washed (25°C, 100 rpm) in about 100 ml of 3X SSC plus 0.1% (w/v) SDS (4 times, 30 min each with fresh solution for each wash). They were then placed in a minimal volume of prehybridization

solution [5X SSC plus 1% w/v each of Ficoll 400 (Pharmacia), polyvinylpyrrolidone (av. M.W. 360,000; Sigma) and bovine serum albumin Fraction V; (Sigma)] for 2 hr at 65°C, 50 rpm. The prehybridization solution was removed, and replaced with the HB14 ³²P-labeled probe that had been saved from the previous hybridization of the library membranes and which had been denatured at 95°C for 5 min. Hybridization was performed at 60°C for 16 hr with shaking at 50 rpm.

Following removal of the labeled probe solution, the membranes were washed 3 times at 25°C (50 rpm, 15 min) in 3X SSC (about 150 ml each wash). They were then washed for 3 hr at 68°C (50 rpm) in 0.25X SSC plus 0.2% SDS (minimal hyb wash solution), and exposed to X-ray film as described above for 1.5 hr at 25°C (no enhancer screens). This exposure revealed very strong hybridization signals to cosmid isolates 22G12, 25A10, 26A5, and 26B10, and a very weak signal with cosmid isolate 8B10. No signal was seen with the negative control (pWE15) colonies, and a very strong signal was seen with positive control membranes (DH5α cells containing the GZ4 isolate of the PCR product) that had been processed concurrently with the experimental samples.

Amplification of a specific genomic fragment of a *tcaB* gene.

Based on the N-terminal amino acid sequence determined for the purified TcaB_i peptide fraction (disclosed here as SEQ ID NO:3) a pool of degenerate oligonucleotides (pool P8F) was synthesized as described for peptide TcaC. The determined amino acid sequence and the corresponding degenerate DNA sequence are given below, where A, C, G, and T are the standard DNA bases, and I represents inosine:

30	Amino Acid	Leu	Phe	Thr	Gln	Thr	Leu	Lys	Glu	Ala	Arg
35	P8F	5' (C/T)TI	TTT	ACI	CA(A/G)	ACI	(C/T)TI	AAA	GAA	GCI	(A/C)G 3'

Another set of degenerate oligonucleotides was synthesized (pool P8.108.3R), representing the complement of the coding strand for the determined amino acid sequence of the TcaB_i-PT108 internal peptide (disclosed herein as SEQ ID NO:20):

Amino Acid	Met	Tyr	Tyr	Ile	Gln	Ala	Gln	Gln
---------------	-----	-----	-----	-----	-----	-----	-----	-----

Codons ATG TA(T/C) TA(T/C) AT(T/C/A) CA(A/G) GC(A/C/G/T) CA(A/G) CA(A/G)
 P8.108.3R 3' AT(A/G) AT(A/G) TA(A/G/T) GT(T/C) CGI GT(T/C) GT 5'
 TAC

5

These oligonucleotides were used as primers for PCR^{*} using HotStart 50 Tubes[™] (Molecular Bio-Products, Inc., San Diego, CA) to amplify a specific DNA fragment from genomic DNA prepared from *Photorhabdus* strain W-14 (see above). A typical reaction (50 μ l) contained (bottom layer) 25 pmol of each primer pool P8F and P8.108.3R, with 2 nmol each of dATP, dCTP, dGTP, and dTTP, in 1X GeneAmp^{*} PCR buffer, and (top layer) 230 ng of genomic template DNA, 8 nmol each of dATP, dCTP, dGTP, and dTTP, and 2.5 units of AmpliTaq^{*} DNA polymerase, in 1X GeneAmp^{*} PCR buffer.

Amplifications were performed by 35 cycles as described for the TcaC peptide. Amplification products were analyzed by electrophoresis through 0.7% w/v SeaKem^{*} LE agarose (FMC) in TEA buffer. A specific product of estimated size 1600 bp was observed.

20

Four such reactions were pooled, and the amplified DNA was extracted from a 1.0% SeaKem^{*} LE gel by Qiaex kit as described for the TcaC peptide. The extracted DNA was used directly as the template for sequence determination (PRISM[™] Sequencing Kit) using the P8F and P8.108.3R primer pools. Each reaction contained about 100 ng template DNA and 25 pmol of one primer pool, and was processed according to standard protocols as described for the TcaC peptide. An analysis of the sequence derived from extension of the P8F primers revealed the short DNA sequence (and encoded amino acid sequence):

30

GAT GCA TTG NTT GCT

Asp Ala Leu (Val) Ala

which corresponds to a portion of the N-terminal peptide sequence disclosed as SEQ ID NO:3 (TcaB_i).

35

Labeling of a TcaB_i-peptide gene-specific probe.

Approximately 50 ng of gel-purified TcaB_i DNA fragment was labeled with ³²P-dCTP as described above, and nonincorporated radioisotopes were removed by passage through a NICK Column^{*} (Pharmacia). The specific activity of the labelled DNA was determined to be 6 x 10⁹ dpm/ μ g. This labeled DNA was used to

probe colony membranes prepared from members of the genomic library that had hybridized to the TcaC-peptide specific probe.

The membranes containing the 12 colonies identified in the TcaC-probe library screen (see above) were stripped of
5 radioactive TcaC-specific label by boiling twice for approximately 30 min each time in 1 liter of 0.1X SSC plus 0.1 % SDS. Removal of radiolabel was checked with a 6 hr film exposure. The stripped membranes were then incubated with the TcaB_i peptide-specific probe prepared above. The labeled DNA was
10 denatured by boiling for 10 min, and then added to the filters that had been incubated for 1 hr in 100 ml of "minimal hyb" solution at 60°C. After overnight hybridization at this temperature, the probe solution was removed, and the filters were washed as follows (all in 0.3X SSC plus 0.1% SDS): once for 5 min
15 at 25°C, once for 1 hr at 60°C in fresh solution, and once for 1 hr at 63°C in fresh solution. After 1.5 hr exposure to X-ray film by standard procedures, 4 strongly-hybridizing colonies were observed. These were, as with the TcaC-specific probe, isolates 22G12, 25A10, 26A5, and 26B10.

20 The same TcaB_iprobe solution was diluted with an equal volume (about 100 ml) of "minimal hyb" solution, and then used to screen the membranes containing the 800 members of the genomic library. After hybridization, washing, and exposure to X-ray film as described above, only the four cosmid clones 22G12,
25 25A10, 26A5, and 26B10, were found to hybridize strongly to this probe.

ISOLATION OF SUBCLONES CONTAINING GENES ENCODING TcaC AND TcaB_i PEPTIDES, AND DETERMINATION OF DNA BASE SEQUENCE THEREOF:

30 Three hybridization-positive cosmids in strain XL1 Blue MR were grown with shaking overnight (200 rpm) at 30°C in 100 ml TB-Amp₁₀₀. After harvesting the cells by centrifugation, cosmid DNA was prepared using a commercially available kit (BIGprep™, 5 Prime 3 Prime, Inc., Boulder, CO), following the manufacturer's
35 protocols. Only one cosmid, 26A5, was successfully isolated by this procedure. When digested with restriction enzyme EcoR 1 (NEB) and analyzed by gel electrophoresis, fragments of approximate sizes 14, 10, 8 (vector), 5, 3.3, 2.9, and 1.5 kbp were detected. A second attempt to isolate cosmid DNA from the
40 same three strains (8 ml cultures; TB-Amp₁₀₀, 30°C) utilized a

boiling miniprep method (Evans G. and G. Wahl., 1987, "Cosmid vectors for genomic walking and rapid restriction mapping." in Guide to Molecular Cloning Techniques. Meth. Enzymology, vol. 152, S. Berger and A. Kimmel, eds., pgs. 604-610). Only one
5 cosmid, 25A10, was successfully isolated by this method. When digested with restriction enzyme EcoR 1 (NEB) and analyzed by gel electrophoresis, this cosmid showed a fragmentation pattern identical to that previously seen with cosmid 26A5.

A 0.15 µg sample of 26A5 cosmid DNA was used to transform 50
10 ml of *E. coli* DH5α cells (Gibco BRL), by the supplier's protocols. A single colony isolate of that strain was inoculated into 4 ml of TB-Amp₁₀₀, and grown for 8 hr at 37°C. Chloramphenicol was added to a final concentration of 225 µg/ml, incubation was continued for another 24 hr, then cells were
15 harvested by centrifugation and frozen at -20°C. Isolation of the 26A5 cosmid DNA was by a standard alkaline lysis miniprep (Maniatis et al., op. cit., p. 382), modified by increasing all volumes by 50% and with stirring or gentle mixing, rather than vortexing, at every step. After washing the DNA pellet in 70%
20 ethanol, it was dissolved in TE containing 25 µg/ml ribonuclease A (Boehringer Mannheim).

Identification of EcoR 1 fragments hybridizing to GZ4-
derived and TcaB₁- probes. Approximately 0.4 µg of cosmid 25A10
25 (from XL1 Blue MR cells) and about 0.5 µg of cosmid 26A5 (from chloramphenicol-amplified DH5α cells) were each digested with about 15 units of EcoR 1 (NEB) for 85 min, frozen overnight, then heated at 65°C for five min, and electrophoresed in a 0.7% agarose gel (Seakem® LE, 1X TEA, 80 volts, 90 min). The DNA was
30 stained with ethidium bromide as described above, and photographed under ultraviolet light. The EcoR 1 digest of cosmid 25A10 was a complete digestion, but the sample of cosmid 26A5 was only partially digested under these conditions. The agarose gel containing the DNA fragments was subjected to
35 depurination, denaturation and neutralization, followed by Southern blotting onto a Magna NT nylon membrane, using a high salt (20X SSC) protocol, all as described in section 2.9 of Ausubel et al. (CPMB, op. cit.). The transferred DNA was then UV-crosslinked to the nylon membrane as before.

An TcaC-peptide specific DNA fragment corresponding to the insert of plasmid isolate GZ4 was amplified by PCR^{*} in a 100 ml reaction volume as described previously above. The amplification products from three such reactions were pooled and were extracted from a 1% GTG^{*} agarose gel by Qiaex kit, as described above, and quantitated by fluorometry. The gel-purified DNA (100 ng) was labeled with ³²P-dCTP using the High Prime Labeling Mix (Boehringer Mannheim) as described above, to a specific activity of 6.34×10^8 dpm/ μ g.

10 The ³²P-labeled GZ4 probe was boiled 10 min, then added to "minimal hyb" buffer (at 1 ng/ml), and the Southern blot membrane containing the digested cosmid DNA fragments was added, and incubated for 4 hr at 60°C with gentle shaking at 50 rpm. The membrane was then washed 3 times at 25°C for about 5 min each (minimal hyb wash solution), followed by two washes for 10 min each at 60°C. The blot was exposed to film (with enhancer screens) for about 30 min at -70°C. The GZ4 probe hybridized strongly to the 5.0 kbp (apparent size) EcoR 1 fragment of both these two cosmids, 26A5 and 25A10.

20 The membrane was stripped of radioactivity by boiling for about 30 min in 0.1X SSC plus 0.1 % SDS, and absence of radiolabel was checked by exposure to film. It was then hybridized at 60°C for 3.5 hours with the (denatured) TcaB_i probe in "minimal hyb" buffer previously used for screening the colony membranes (above), washed as described previously, and exposed to film for 40 min at -70°C with two enhancer screens. With both cosmids, the TcaB_i probe hybridized lightly with the about 5.0 kbp EcoR 1 fragment, and strongly with a fragment of approximately 2.9 kbp.

30 The sample of cosmid 26A5 DNA previously described, (from DH5 α cells) was used as the source of DNA from which to subclone the bands of interest. This DNA (2.5 μ g) was digested with about 3 units of EcoR 1 (NEB) in a total volume of 30 μ l for 1.5 hr, to give a partial digest, as confirmed by gel electrophoresis. Ten μ g of pBC KS (+) DNA (Stratagene) were digested for 1.5 hr with 20 units of EcoR 1 in a total volume of 20 μ l, leading to total digestion as confirmed by electrophoresis. Both EcoR 1-cut DNA preparations were diluted to 50 μ l with water, to each an equal volume of PCI was added, the suspension was gently mixed, spun in

a microcentrifuge and the aqueous supernatant was collected. DNA was precipitated by 150 μ l ethanol, and the mixture was placed at -20°C overnight. Following centrifugation and drying, the EcoR 1-digested pBC KS (+) was dissolved in 100 μ l TE; the partially
5 digested 26A5 was dissolved in 20 μ l TE. DNA recovery was checked by fluorometry.

In separate reactions, approximately 60 ng of EcoR 1-digested pBC KS(+) DNA was ligated with approximately 180 ng or 270 ng of partially digested cosmid 26A5 DNA. Ligations were
10 carried out in a volume of 20 μ l at 15°C for 5 hr, using T4 ligase and buffer from New England BioLabs. The ligation mixture, diluted to 100 μ l with sterile TE, was used to transform frozen, competent DH5 α cells (Gibco BRL) according to the supplier's instructions. Varying amounts (25-200 μ l) of the
15 transformed cells were plated on freshly prepared solid LB-Cam_rs medium with 1 mM IPTG and 50 mg/l X-gal. Plates were incubated at 37°C about 20 hr, then chilled in the dark for approximately 3 hr to intensify color for insert selection. White colonies were picked onto patch plates of the same composition and incubated
20 overnight at 37°C.

Two colony lifts of each of the selected patch plates were prepared as follows. After picking white colonies to fresh plates, round Magna NT nylon membranes were pressed onto the patch plates, the membrane was lifted off, and subjected to
25 denaturation, neutralization and UV crosslinking as described above for the library colony membranes. The crosslinked colony lifts were vigorously washed, including gently wiping off the excess cell debris with a tissue. One set was hybridized with the GZ4(TcaC) probe solution described earlier, and the other set
30 was hybridized with the TcaB_i probe solution described earlier, according to the 'minimal hyb' protocol, followed by washing and film exposure as described for the library colony membranes.

Colonies showing hybridization signals either only with the GZ4 probe, with both GZ4 and TcaB_i probes, or only with the TcaB_i
35 probe, were selected for further work and cells were streaked for single colony isolation onto LB-Cam_rs media with IPTG and X-gal as before. Approximately 35 single colonies, from 16 different isolates, were picked into liquid LB-Cam_rs media and grown

overnight at 37°C; the cells were collected by centrifugation and plasmid DNA was isolated by a standard alkaline lysis miniprep according to Maniatis et al. (op. cit. p. 368). DNA pellets were dissolved in TE + 25 µg/ml ribonuclease A and DNA concentration was determined by fluorometry. The EcoR 1 digestion pattern was analyzed by gel electrophoresis. The following isolates were picked as useful. Isolate A17.2 contains religated pBC KS(+) only and was used for a (negative) control. Isolates D38.3 and C44.1 each contain only the 2.9 kbp, TcaB_i -hybridizing EcoR 1 fragment inserted into pBC KS(+). These plasmids, named pDAB2000 and pDAB2001, respectively, are illustrated in Fig. 2.

Isolate A35.3 contains only the approximately 5 kbp, GZ4)-hybridizing EcoR 1 fragment, inserted into pBC KS(+). This plasmid was named pDAB2002 (also Fig. 2). These isolates provided templates for DNA sequencing.

Plasmids pDAB2000 and pDAB2001 were prepared using the BIGprep™ kit as before. Cultures (30 ml) were grown overnight in TB-Cam₃ to an OD₆₀₀ of 2, then plasmid was isolated according to the manufacturer's directions. DNA pellets were redissolved in 100 µl TE each, and sample integrity was checked by EcoR 1 digestion and gel electrophoretic analysis.

Sequencing reactions were run in duplicate, with one replicate using as template pDAB2000 DNA, and the other replicate using as template pDAB2001 DNA. The reactions were carried out using the dideoxy dye terminator cycle sequencing method, as described above for the sequencing of the GZ4/HB14 DNAs. Initial sequencing runs utilized as primers the LacZ and T7 primers described above, plus primers based on the determined sequence of the TcaB_i PCR amplification product (TH1 = ATTGCAGACTGCCAATCGCTTCGG, TH12 = GAGAGTATCCAGACCGCGGATGATCTG).

After alignment and editing of each sequencing output, each was truncated to between 250 to 350 bases, depending on the integrity of the chromatographic data as interpreted by the Perkin Elmer Applied Biosystems Division SeqEd 675 software. Subsequent sequencing "steps" were made by selecting appropriate sequence for new primers. With a few exceptions, primers (synthesized as described above) were 24 bases in length with a 50% G+C composition. Sequencing by this method was carried out on both strands of the approximately 2.9 kbp EcoR 1 fragment.

To further serve as template for DNA sequencing, plasmid DNA from isolate pDAB2002 was prepared by BIGprep™ kit. Sequencing reactions were performed and analyzed as described above.

Initially, a T3 primer (pBS SK (+) bases 774-796:

5 CGCGCAATTAACCCTCACTAAAG) and a T7 primer (pBS KS (+) bases 621-643: GCGCGTAATACGACTCACTATAG) were used to prime the sequencing reactions from the flanking vector sequences, reading into the insert DNA. Another set of primers, (GZ4F:

10 GTATCGATTACAACGCTGCTCACTTCCC; TH13: GGGAAGTGACAGCGTTGTAATCGATAC; TH14: ATGTTGGGTGCGTCGGCTAATGGACATAAC; and LW1-204: GGGAAGTGACAGCGTTGTAATCGATAC) was made to prime from internal sequences, which were determined previously by degenerate oligonucleotide-mediated sequencing of subcloned TcaC-peptide PCR products. From the data generated during the initial rounds of
15 sequencing, new sets of primers were designed and used to walk the entire length of the ~5 kbp fragment. A total of 55 oligo primers was used, enabling the identification of 4832 total bp of contiguous sequence.

When the DNA sequence of the EcoR 1 fragment insert of
20 pDAB2002 is combined with part of the determined sequence of the pDAB2000/pDAB2001 isolates, a total contiguous sequence of 6005 bp was generated (disclosed herein as SEQ ID NO:25). When long open reading frames were translated into the corresponding amino acids, the sequence clearly shows the TcaB_i N-terminal peptide
25 (disclosed as SEQ ID NO:3), encoded by bases 19-75, immediately following a methionine residue (start of translation). Upstream lies a potential ribosome binding site (bases 1-9), and downstream, at bases 166-228 is encoded the TcaB_i-PT158 internal peptide (disclosed herein as SEQ ID NO:19). Further downstream,
30 in the same reading frame, at bases 1738-1773, exists a sequence encoding the TcaB_i-PT108 internal peptide (disclosed herein as SEQ ID NO:20). Also in the same reading frame, at bases 1897-1923, is encoded the TcaB_{ii} N-terminal peptide (disclosed herein as SEQ ID NO:5), and the reading frame continues uninterrupted to
35 a translation termination codon at nucleotides 3586-3588.

The lack of an in-frame stop codon between the end of the sequence encoding TcaB_i-PT108 and the start of the TcaB_{ii} encoding region, and the lack of a discernible ribosome binding site immediately upstream of the TcaB_{ii} coding region, indicate that

peptides TcaB_{ii} and TcaB_i are encoded by a single open reading frame of 3567 bp beginning at base pair 16 in SEQ ID NO:25), and are most likely derived from a single primary gene product of 1189 amino acids (131,586 Daltons; disclosed herein as SEQ ID NO:26) by post-translational cleavage. If the amino acid immediately preceding the TcaB_{ii} N-terminal peptide represents the C-terminal amino acid of peptide TcaB_i, then the predicted mass of TcaB_{ii} (627 amino acids) is 70,814 Daltons (disclosed herein as SEQ ID NO:28), somewhat higher than the size observed by SDS-PAGE (68 kDa). This peptide would be encoded by a contiguous stretch of 1881 base pairs (disclosed herein as SEQ ID NO:27). It is thought that the native C-terminus of TcaB_i lies somewhat closer to the C-terminus of TcaB_i-PT108. The molecular mass of PT108 [3.438 kDa; determined during N-terminal amino acid sequence analysis of this peptide] predicts a size of 30 amino acids. Using the size of this peptide to designate the C-terminus of the TcaB_i coding region [Glu at position 604 of SEQ ID NO:28], the derived size of TcaB_i is determined to be 604 amino acids or 68,463 Daltons, more in agreement with experimental observations.

Translation of the TcaB_{ii} peptide coding region of 1686 base pairs (disclosed herein as SEQ ID NO:29) yields a protein of 562 amino acids (disclosed herein as SEQ ID NO:30) with predicted mass of 60,789 Daltons, which corresponds well with the observed 61 kDa.

A potential ribosome binding site (bases 3633-3638) is found 48 bp downstream of the stop codon for the *tcaB* open reading frame. At bases 3645-3677 is found a sequence encoding the N-terminus of peptide TcaC, (disclosed as SEQ ID NO.2). The open reading frame initiated by this N-terminal peptide continues uninterrupted to base 6005 (2361 base pairs, disclosed herein as the first 2361 base pairs of SEQ ID NO.31). A gene (*tcaC*) encoding the entire TcaC peptide, (apparent size ~165 kDa; ~1500 amino acids), would comprise about 4500 bp.

Another isolate containing cloned EcoR 1 fragments of cosmid 26A5, E20.6, was also identified by its homology to the previously mentioned GZ4 and TcaB_i probes. Agarose gel analysis of EcoR 1 digests of the DNA of the plasmid harbored by this strain (pDAB2004, Fig. 2), revealed insert fragments of estimated

sizes 2.9, 5, and 3.3 kbp. DNA sequence analysis initiated from primers designed from the sequence of plasmid pDAB2002 revealed that the 3.3 kbp EcoR 1 fragment of pDAB2004 lies adjacent to the 5 kbp EcoR 1 fragment represented in pDAB2002. The 2361 base pair open reading frame discovered in pDAB2002 continues uninterrupted for another 2094 bases in pDAB2004 [disclosed herein as base pairs 2362 to 4458 of SEQ ID NO:31]. DNA sequence analysis using the parent cosmid 26A5 DNA as template confirmed the continuity of the open reading frame. Altogether, the open reading frame (TcaC SEQ ID NO:31) comprises 4455 base pairs, and encodes a protein (TcaC) of 1485 amino acids [disclosed herein as SEQ ID NO:32]. The calculated molecular size of 166,214 Daltons is consistent with the estimated size of the TcaC peptide (165 kDa), and the derived amino acid sequence matches exactly that disclosed for the TcaC N-terminal sequence [SEQ ID NO:2].

The lack of an amino acid sequence corresponding to SEQ ID NO:17; used to design the degenerate oligonucleotide primer pool in the discovered sequence indicates that the generation of the PCR® products found in isolates GZ4 and HB14, which were used as probes in the initial library screen, were fortuitously generated by reverse-strand priming by one of the primers in the degenerate pool. Further, the derived protein sequence does not include the internal fragment disclosed herein as SEQ ID NO:18. These sequences reveal that plasmid pDAB2004 contains the complete coding region for the TcaC peptide.

Example 9

Screening of the *Photorhabdus* genomic library for genes encoding the TcbA_{ij} peptide

30

This example describes a method used to identify DNA clones that contain the TcbA_{ij} peptide-encoding genes, the isolation of the gene, and the determination of its partial DNA base sequence.

Primers and PCR reactions

The TcbA_{ij} polypeptide of the insect active preparation is ~206 kDa. The amino acid sequence of the N-terminus of this peptide is disclosed as SEQ ID NO:1. Four pools of degenerate oligonucleotide primers ("Forward primers": TH-4, TH-5, TH-6, and

TH-7) were synthesized to encode a portion of this amino acid sequence, as described in Example 8, and are shown below.

Table 11

5	Amino Acid	Phe	Ile	Gln	Gly	Tyr	Ser	Asp	Leu	Phe
	TH-4	5'-TT(T/C)	ATI	CA(A/G)	GGI	TA(T/C)	TCI	GA(T/C)	CTI	TT-3'
	TH-5	5'-TT(T/C)	ATI	CA(A/G)	GGI	TA(T/C)	AG(T/C)	GA(T/C)	CTI	TT-3'
	TH-6	5'-TT(T/C)	ATI	CA(A/G)	GGI	TA(T/C)	TCI	GA(T/C)	TT(A/G)	TT-3'
10	TH-7	5'-TT(T/C)	ATI	CA(A/G)	GGI	TA(T/C)	AG(T/C)	GA(T/C)	TT(A/G)	TT-3'

In addition, a primary ("a") and a secondary ("b") sequence of an internal peptide preparation (TcbA₁₁-PT81) have been determined and are disclosed herein as SEQ ID No:23 and SEQ ID No:24, respectively. Four pools of degenerate oligonucleotides ("Reverse Primers": TH-8, TH-9, TH-10 and TH-11) were similarly designed and synthesized to encode the reverse complement of sequences that encode a portion of the peptide of SEQ ID NO:23, as shown below.

Table 12

Amino Acid	Thr	Tyr	Leu	Thr	Ser	Phe	Glu	Gln	Val	Ala	Asn
TH-8	3'TGI	AT(A/G)	GAI	TGI	AGI	AA(A/G)	CT(T/C)	GT(T/C)	CAI	CGI	TT(G/A)-5'
TH-9	3'TGI	AT(A/G)	TT(A/G)	TGI	AGI	AA(A/G)	CT(T/C)	GT(T/C)	CAI	CGI	TT(G/A)-5'
TH-10	3'TGI	AT(A/G)	GAI	TGI	TC(G/A)	AA(A/G)	CT(T/C)	GT(T/C)	CAI	CGI	TT(G/A)-5'
TH-11	3'TGI	AT(A/G)	TT(A/G)	TGI	TC(G/A)	AA(A/G)	CT(T/C)	GT(T/C)	CAI	CGI	TT(G/A)-5'

Sets of these primers were used in PCR[®] reactions to amplify TcbAii- encoding gene fragments from the genomic *Photorhabdus luminescens* W-14 DNA prepared in Example 6. All PCR[®] reactions were run with the "Hot Start" technique using AmpliWax[™] gems and other Perkin Elmer reagents and protocols. Typically, a mixture (total volume 11 µl) of MgCl₂, dNTP's, 10X GeneAmp[®] PCR Buffer II, and the primers were added to tubes containing a single wax bead. [10X GeneAmp[®] PCR Buffer II is composed of 100 mM Tris-HCl, pH 8.3; and 500 mM KCl.] The tubes were heated to 80°C for 2 minutes and allowed to cool. To the top of the wax seals, a solution containing 10X GeneAmp[®] PCR Buffer II, DNA template, and AmpliTaq[®] DNA polymerase were added. Following melting of the wax seal and mixing of components by thermal cycling, final reaction conditions (volume of 50 µl) were: 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 2.5 mM MgCl₂; 200 µM each in dATP, dCTP, dGTP, dTTP; 1.25 mM in a single Forward primer pool; 1.25 µM in a single Reverse primer pool, 1.25 units of AmpliTaq[®] DNA polymerase, and 170 ng of template DNA.

The reactions were placed in a thermocycler (as in Example 8) and run with the following program:

Table 13

Temperature	Time	Cycle Repetition
94°C	2 minutes	1X
94°C	15 seconds	30X
55-65°C	30 seconds	
72°C	1 minute	
72°C	7 minutes	1X
15°C	Constant	

A series of amplifications was run at three different annealing temperatures (55°, 60°, 65° C) using the degenerate primer pools. Reactions with annealing at 65°C had no amplification products visible following agarose gel electrophoresis. Reactions having a 60°C annealing regime and containing primers TH-5+TH-10 produced an amplification product that had a mobility corresponding to 2.9 kbp. A lesser amount of the 2.9 kbp product was produced under these conditions with primers TH-7+TH-10. When reactions were annealed at 55°C, these primer pairs produced more of the 2.9 kbp product, and this product was also produced by primer pairs TH-5+TH-8 and TH-5+TH-11. Additional very faint 2.9 kbp bands were seen in lanes containing amplification products from primer pairs TH-7 plus TH-8, TH-9, TH-10, or TH-11.

To obtain sufficient PCR amplification product for cloning and DNA sequence determination, 10 separate PCR reactions were set up using the primers TH-5+TH-10, and were run using the above conditions with a 55°C annealing temperature. All reactions were pooled and the 2.9 kbp product was purified by Qiaex extraction from an agarose gel as described above.

Additional sequences determined for TcbA_{ij} internal peptides are disclosed herein as SEQ ID NO:21 and SEQ ID NO:22. As before, degenerate oligonucleotides (Reverse primers TH-17 and TH-18) were made corresponding to the reverse complement of sequences that encode a portion of the amino acid sequence of these peptides.

Table 14

From SEQ ID NO:21

Amino Acid	Met	Glu	Thr	Gln	Asn	Ile	Gln	Glu	Pro
TH-17	3'-TAC	CTT/C	TGI	GTT/C	TTA/G	TAI	GTT/C	GTT/C	GG-5'

Table 15

From SEQ ID NO:22

Amino Acid	Asn	Pro	Ile	Asn	Ile	Asn	Thr	Gly	Ile	Asp
TH-18	3'-TT(A/G)	GGI	TAI	TT(A/G)	TAI	TT(A?G)	TGI	CCI	TAI	CT(A/G)-5'

Degenerate oligonucleotides TH-18 and TH-17 were used in an amplification experiment with *Photobacterium luminescens* W-14 DNA as template and primers TH-4, TH-5, TH-6, or TH-7 as the 5'-(Forward) primers. These reactions amplified products of approximately 4 kbp and 4.5 kbp, respectively. These DNAs were transferred from agarose gels to nylon membranes and hybridized with a ³²P-labeled probe (as described above) prepared from the 2.9 kbp product amplified by the TH-5+TH10 primer pair. Both the 4 kbp and the 4.5 kbp amplification products hybridized strongly to the 2.9 kbp probe. These results were used to construct a map ordering the TcbA_{ii} internal peptide sequences as shown in Fig. 3. Approximate distances between the primers are shown in nucleotides in Fig. 3.

15 DNA Sequence of the 2.9 kbp TcbA_{ii}-encoding fragment

Approximately 200 ng of the purified 2.9 kbp fragment (prepared above) was precipitated with ethanol and dissolved in 17 ml of water. One-half of this was used as sequencing template with 25 pmol of the TH-5 pool as primers, the other half was used as template for TH-10 priming. Sequencing reactions were as given in Example 8. No reliable sequence was produced using the TH-10 primer pool; however, reactions with TH-5 primer pool produced the sequence disclosed below:.

```

1   AATCGTGTG ATCCCTATGC CGNGCCGGGT TCGGTGGAAT CGATGTCCTC ACCGGGGGTT
25  51   TATTNGAGGG ANTNGTCCCG TGAGGCCAAA AANTGGAATG AAAGAAGTTC AATTINTTAC
    121  CTAGATAAAC GTCGCCCCGN TTTAGAAAGN TTANTGNTCA GCCAGAAAAT TTTGGTTGAG
    181  GAAATTCCAC CGNTGGTTCT CTCTATTGAT TNGGGCCTGG CCGGGTTCTGA ANNAAAACNA
    241  GGAAATNCAC AAGTTGAGGT GATGGNTTTG TNGCNANCTT NTCGTTTAGG TGGGGAGAAA
    301  CCTTNTCANC ACGNTTNTGA AACTGTCCGG GAAATCGTCC ATGANCCTGA NCCAGGNTTN
30  361  CGCCATTGG

```

Based on this sequence, a sequencing primer (TH-21, 5'-CCGGGCGACGTTTATCTAGG-3') was designed to reverse complement bases 120-139, and initiate polymerization towards the 5' end (i.e., TH-5 end) of the gel-purified 2.9 kbp TcbA_{ii}-encoding PCR fragment. The determined sequence is shown below, and is compared to the biochemically determined N-terminal peptide sequence of TcbA_{ii} SEQ ID NO:1.

TcbAii 2.9 kbp PCR fragment Sequence Confirmation

[Underlined amino acids = encoded by degenerate oligonucleotides]

5 SEQ ID NO:1 F I Q G Y S D L F G - - A
 2.9 kbp seq GC ATG CAG GGG TAT AGT GAC CTG TTT GGT AAT CGT GCT
 M Q G Y S D L F G N R A

10 From the homology of the derived amino acid sequence to the
 biochemically determined one, it is clear that the 2.9 kbp PCR
 fragment represents the *TcbA* coding region. This 2.9 kbp
 fragment was then used as a hybridization probe to screen the
Phototrhobdus W-14 genomic library prepared in Example 8 for
 cosmids containing the *TcbAii*-encoding gene.

15

Screening the *Phototrhobdus* cosmid library

The 2.9 kb gel-purified PCR fragment was labeled with ³²P
 using the Boehringer Mannheim High Prime labeling kit as
 described in Example 8. Filters containing remnants of
 20 approximately 800 colonies from the cosmid library were screened
 as described previously (Example 8), and positive clones were
 streaked for isolated colonies and rescreened. Three clones
 (8A11, 25G8, and 26D1) gave positive results through several
 screening and characterization steps. No hybridization of the
 25 *TcbAii*-specific probe was ever observed with any of the four
 cosmids identified in Example 8, and which contain the *tcaB* and
tcaC genes. DNA from cosmids 8A11, 25G8, and 26D1 was digested
 with restriction enzymes Bgl 2, EcoR 1 or Hind 3 (either alone or
 in combination with one another), and the fragments were
 30 separated on an agarose gel and transferred to a nylon membrane
 as described in Example 8. The membrane was hybridized with ³²P-
 labeled probe prepared from the 4.5 kbp fragment (generated by
 amplification of *Phototrhobdus* genomic DNA with primers TH-5+TH-
 17). The patterns generated from cosmid DNAs 8A11 and 26D1 were
 35 identical to those generated with similarly-cut genomic DNA on
 the same membrane. It is concluded that cosmids 8A11 and 26D1
 are accurate representations of the genomic *TcbAii* encoding
 locus. However, cosmid 25G8 has a single Bgl 2 fragment which is
 slightly larger than the genomic DNA. This may result from
 40 positioning of the insert within the vector.

DNA sequence of the tcbA-encoding gene

The membrane hybridization analysis of cosmid 26D1 revealed that the 4.5 kbp probe hybridized to a single large EcoR 1 fragment (greater than 9 kbp). This fragment was gel purified and ligated into the EcoR 1 site of pBC KS (+) as described in Example 8, to generate plasmid pBC-S1/R1. The partial DNA sequence of the insert DNA of this plasmid was determined by "primer walking" from the flanking vector sequence, using procedures described in Example 8. Further sequence was generated by extension from new oligonucleotides designed from the previously determined sequence. When compared to the determined DNA sequence for the tcbA gene identified by other methods (disclosed herein as SEQ ID NO:11 as described in Example 12 below), complete homology was found to nucleotides 1-272, 319-826, 2578-3036, and 3068-3540 (total bases = 1712). It was concluded that both approaches can be used to identify DNA fragments encoding the TcbA_{ii} peptide.

Analysis of the derived amino acid sequence of the tcbA gene.

The sequence of the DNA fragment identified as SEQ ID NO:11 encodes a protein whose derived amino acid sequence is disclosed herein as SEQ ID NO:12. Several features verify the identity of the gene as that encoding the TcbA_{ii} protein. The TcbA_{ii} N-terminal peptide (SEQ ID NO:1; Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala) is encoded as amino acids 88-100. The TcbA_{ii} internal peptide TcbA_{ii}-PT81(a) (SEQ ID NO:23) is encoded as amino acids 1065-1077, and TcbA_{ii}-PT81(b) (SEQ ID NO:24) is encoded as amino acids 1571-1592. Further, the internal peptide TcbA_{ii}-PT56 (SEQ ID NO:22) is encoded as amino acids 1474-1488, and the internal peptide TcbA_{ii}-PT103 (SEQ ID NO:24) is encoded as amino acids 1614-1639. It is obvious that this gene is an authentic clone encoding the TcbA_{ii} peptide as isolated from insecticidal protein preparations of *Photobacterium luminescens* strain W-14.

The protein isolated as peptide TcbA_{ii} is derived from cleavage of a longer peptide. Evidence for this is provided by the fact that the nucleotides encoding the TcbA_{ii} N-terminal peptide SEQ ID NO:1 are preceded by 261 bases (encoding 87 N-terminal-proximal amino acids) of a longer open reading frame (SEQ ID NO:11). This reading frame begins with nucleotides that encode the amino acid sequence Met Gln Asn Ser

Leu, which corresponds to the N-terminal sequence of the large peptide TcbA, and is disclosed herein as SEQ ID NO:16. It is thought that TcbA is the precursor protein for TcbA_{ii}.

5 Relationship of tcbA, tcaB and tcaC genes.

The *tcaB* and *tcaC* genes are closely linked and may be transcribed as a single mRNA (Example 8). The *tcbA* gene is borne on cosmids that apparently do not overlap the ones harboring the *tcaB* and *tcaC* cluster, since the respective genomic library
10 screens identified different cosmids. However, comparison of the amino sequences encoded by the *tcaB* and *tcaC* genes with the *tcbA* gene reveals a substantial degree of homology. The amino acid conservation (Protein Alignment Mode of MacVector™ Sequence
15 Scientific Imaging Systems, Rochester, NY) is shown in Fig. 4. On the score line of each panel in Fig. 4, up carats (^) indicate homology or conservative amino acid changes, and down carats (v) indicate nonhomology.

This analysis shows that the amino acid sequence of the TcbA
20 peptide from residues 1739 to 1894 is highly homologous to amino acids 441 to 603 of the TcaB_i peptide (162 of the total 627 amino acids of P8; SEQ ID NO:28). In addition, the sequence of TcbA amino acids 1932 to 2459 is highly homologous to amino acids 12 to 531 of peptide TcaB_{ii} (520 of the total 562 amino acids; SEQ
25 ID NO:30). Considering that the TcbA peptide (SEQ ID NO:12) comprises 2505 amino acids, a total of 684 amino acids (27%) at the C-proximal end of it is homologous to the TcaB_i or TcaB_{ii} peptides, and the homologies are arranged colinear to the arrangement of the putative TcaB preprotein (SEQ ID NO:26). A
30 sizeable gap in the TcbA homology coincides with the junction between the TcaB_i and TcaB_{ii} portions of the TcaB preprotein. Clearly the TcbA and TcaB gene products are evolutionarily related, and it is proposed that they share some common
35 function(s) in *Photorhabdus*.

Example 10Characterization of zinc-metalloproteases in *Phototrhhabdus* Broth:
Protease Inhibition, Classification, and Purification

5 Protease Inhibition and Classification Assays: Protease assays were performed using FITC-casein dissolved in water as substrate (0.08% final assay concentration). Proteolysis reactions were performed at 25°C for 1 h in the appropriate buffer with 25 µl of *Phototrhhabdus* broth (150 µl total reaction
10 volume). Samples were also assayed in the presence and absence of dithiothreitol. After incubation, an equal volume of 12% trichloroacetic acid was added to precipitate undigested protein. Following precipitation for 0.5 h and subsequent centrifugation, 100 µl of the supernatant was placed into a 96-well microtiter
15 plate and the pH of the solution was adjusted by addition of an equal volume of 4N NaOH. Proteolysis was then quantitated using a Fluoroskan II fluorometric plate reader at excitation and emission wavelengths of 485 and 538 nm, respectively. Protease activity was tested over a range from pH 5.0-10.0 in 0.5 units
20 increments. The following buffers were used at 50 mM final concentration: sodium acetate (pH 5.0 - 6.5); Tris-HCL (pH 7.0 - 8.0); and bis-Tris propane (pH 8.5-10.0). To identify the class of protease(s) observed, crude broth was treated with a variety of protease inhibitors (0.5 µg/µl final concentration) and then
25 examined for protease activity at pH 8.0 using the substrate described above. The protease inhibitors used included E-64 (L-trans-expoxysaccinylleucylamido(4-, -guanidino)-butane), 3,4 dichloroisocoumarin, Leupeptin, pepstatin, amastatin, ethylenediaminetetraacetic acid (EDTA) and 1,10 phenanthroline.

30 Protease assays performed over a pH range revealed that indeed protease(s) were present which exhibited maximal activity at ~ pH 8.0 (Table 16). Addition of DTT did not have any effect on protease activity. Crude broth was then treated with a variety of protease inhibitors (Table 17). Treatment of crude
35 broth with the inhibitors described above revealed that 1,10 phenanthroline caused complete inhibition of all protease activity when added at a final concentration of 50 µg, with the IC₅₀ = 5 µg in 100 µl of a 2 mg/ml crude broth solution. These data indicate that the most abundant protease(s) found in the

Photorhabdus broth are from the zinc-metalloprotease class of enzymes.

Table 16
5 Effect of pH on the protease activity found in a Day 1 production of *Photorhabdus luminescens* (strain W-14).

	pH	Flu. Units ^a Activity ^b		Percent
10	5.0	3013 ±	78	17
	5.5	7994 ±	448	45
15	6.0	12965 ±	483	74
	6.5	14390 ±	1291	82
	7.0	14386 ±	1287	82
20	7.5	14135 ±	198	80
	8.0	17582 ±	831	100
25	8.5	16183 ±	953	92
	9.0	16795 ±	760	96
	9.5	16279 ±	1022	93
30	10.0	15225 ±	210	87

a Flu. Units = Fluorescence Units (Maximum = ~28,000; background = ~2200).

b. Percent activity relative to the maximum at pH 8.0

35

Table 17
Effect of different protease inhibitors on the protease activity
at pH 8 found in a Day 1 production of *Photobacterium luminescens*
(strain W-14).

5

Inhibitor	Corrected Flu. Units ^a	Percent Inhibition ^b
Control	13053	0
E-64	14259	0
10 1,10 Phenanthroline ^c	15	99
3,4 Dichloroisocoumarin ^d	7956	39
Leupeptin	13074	0
Pepstatin ^c	13441	0
Amastatin	12474	4
15 DMSO Control	12005	8
Methanol Control	12125	7

a. Corrected Flu. Units = Fluorescence Units - background(2200 flu. units).

20 b. Percent Inhibition relative to protease activity at pH 8.0.

c. Inhibitors were dissolved in methanol.

d. Inhibitors were dissolved in DMSO.

The isolation of a zinc-metalloprotease was performed by
25 applying dialyzed 10-80% ammonium sulfate pellet to a Q Sepharose
column equilibrated at 50 mM Na₂PO₄, pH 7.0 as described in
Example 5 for *Photobacterium* toxin. After extensive washing, a 0
to 0.5 M NaCl gradient was used to elute toxin protein. The
majority of biological activity and protein was eluted from 0.15
30 - 0.45 M NaCl. However, it was observed that the majority of
proteolytic activity was present in the 0.25-0.35 M NaCl fraction
with some activity in the 0.15-0.25 M NaCl fraction. SDS PAGE
analysis of the 0.25-0.35 M NaCl fraction showed a major peptide
band of approximately 60 kDa. The 0.15-0.25 M NaCl fraction
35 contained a similar 60 kDa band but at lower relative protein
concentration. Subsequent gel filtration of this fraction using
a Superose 12 HR 16/50 column resulted in a major peak migrating
at 57.5 kDa that contained a predominant (> 90% of total stained
protein) 58.5 kDa band by SDS PAGE analysis. Additional analysis
40 of this fraction using various protease inhibitors as described
above determined that the protease was a zinc-metalloprotease.
Nearly all of the protease activity present in *Photobacterium* broth
at day 1 of fermentation corresponded to the ~58 kDa zinc-
metalloprotease.

45 In yet a second isolation of zinc-metalloprotease(s), W-14
Photobacterium broth grown for three days was taken and protease

activity was visualized using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) laced with gelatin as described in Schmidt, T.M., Bleakley, B. and Nealson, K.M. 1988. SDS running gels (5.5 x 8 cm) were made with 12.5 % polyacrylamide (40% stock solution of acrylamide/bis-acrylamide; Sigma Chemical Co., St. Louis, MO) into which 0.1% gelatin final concentration (Biorad EIA grade reagent; Richmond CA) was incorporated upon dissolving in water. SDS-stacking gels (1.0 x 8 cm) were made with 5% polyacrylamide, also laced with 0.1% gelatin. Typically, 2.5 µg of protein to be tested was diluted in 0.03 ml of SDS-PAGE loading buffer without dithiothreitol (DTT) and loaded onto the gel. Proteins were electrophoresed in SDS running buffer (Laemmli, U.K. 1970. Nature 227, 680) at 0° C and at 8 mA. After electrophoresis was complete, the gel was washed for 2 h in 2.5% (v/v) Triton X-100. Gels were then incubated for 1 h at 37 °C in 0.1 M glycine (pH 8.0). After incubation, gels were fixed and stained overnight with 0.1% amido black in methanol-acetic acid- water (30:10:60, vol./vol./vol.; Sigma Chemical Co.). Protease activity was visualized as light areas against a dark, amido black stained background due to proteolysis and subsequent diffusion of incorporated gelatin. At least three distinct bands produced by proteolytic activity at 58-, 41-, and 38 kDa were observed.

Activity assays of the different proteases in W-14 day three culture broth were performed using FITC-casein dissolved in water as substrate (0.02% final assay concentration). Proteolysis experiments were performed at 37 °C for 0-0.5 h in 0.1M Tris-HCl (pH 8.0) with different protein fractions in a total volume of 0.15 ml. Reactions were terminated by addition of an equal volume of 12% trichloroacetic acid (TCA) dissolved in water. After incubation at room temperature for 0.25 h, samples were centrifuged at 10,000 x g for 0.25 h and 0.10 ml aliquots were removed and placed into 96-well microtiter plates. The solution was then neutralized by the addition of an equal volume of 2 N sodium hydroxide, followed by quantitation using a Fluoroskan II fluorometric plate reader with excitation and emission wavelengths of 485 and 538 nm, respectively. Activity measurements were performed using FITC-Casein with different protease concentrations at 37° C for 0-10 min. A unit of

activity was arbitrarily defined as the amount of enzyme needed to produce 1000 fluorescent units/min and specific activity was defined as units/mg of protease.

Inhibition studies were performed using two zinc-
5 metalloprotease inhibitors; 1,10 phenanthroline and N-(α -
rhamnopyranosyloxyhydroxyphosphinyl)-Leu-Trp(phosphoramidon) with
stock solutions of the inhibitors dissolved in 100% ethanol and
water, respectively. Stock concentrations were typically 10
mg/ml and 5 mg/ml for 1,10 phenanthroline and phosphoramidon,
10 respectively, with final concentrations of inhibitor at 0.5-1.0
mg/ml per reaction. Treatment of three day W-14 crude broth with
1,10 phenanthroline, an inhibitor of all zinc metalloproteases,
resulted in complete elimination of all protease activity while
treatment with phosphoramidon, an inhibitor of thermolysin-like
15 proteases (Weaver, L.H., Kester, W.R., and Matthews, B.W. 1977.
J. Mol. Biol. 114, 119-132), resulted in ~56% reduction of
protease activity. The residual proteolytic activity could not
be further reduced with additional phosphoramidon.

The proteases of three day W-14 *Photorhabdus* broth were
20 purified as follows: 4.0 liters of broth were concentrated using
an Amicon spiral ultra filtration cartridge Type SLY100 attached
to an Amicon M-12 filtration device. The flow-through material
having native proteins less than 100 kDa in size (3.8 L) was
concentrated to 0.375 L using an Amicon spiral ultra filtration
25 cartridge Type SLY10 attached to an Amicon M-12 filtration
device. The retentate material contained proteins ranging in
size from 10-100 kDa. This material was loaded onto a Pharmacia
HR16/10 column which had been packed with PerSeptive Biosystem
(Framington, MA) Poros® 50 HQ strong anion exchange packing that
30 had been equilibrated in 10 mM sodium phosphate buffer (pH 7.0).
Proteins were loaded on the column at a flow rate of 5 ml/min,
followed by washing unbound protein with buffer until A₂₈₀ =
0.00. Afterwards, proteins were eluted using a NaCl gradient of
0-1.0 M NaCl in 40 min at a flow rate of 7.5 ml/min. Fractions
35 were assayed for protease activity, supra., and active fractions
were pooled. Proteolytically active fractions were diluted with
50% (v/v) 10 mM sodium phosphate buffer (pH 7.0) and loaded onto
a Pharmacia HR 10/10 Mono Q column equilibrated in 10 mM sodium
phosphate. After washing the column with buffer until A₂₈₀ =

0.00, proteins were eluted using a NaCl gradient of 0-0.5 M NaCl for 1 h at a flow rate of 2.0 ml/min. Fractions were assayed for protease activity. Those fractions having the greatest amount of phosphoramidon-sensitive protease activity, the phosphoramidon sensitive activity being due to the 41/38 kDa protease, *infra.*, were pooled. These fractions were found to elute at a range of 0.15-0.25 M NaCl. Fractions containing a predominance of phosphoramidon-insensitive protease activity, the 58 kDa protease, were also pooled. These fractions were found to elute at a range of 0.25-0.35 M NaCl. The phosphoramidon-sensitive protease fractions were then concentrated to a final volume of 0.75 ml using a Millipore Ultrafree®-15 centrifugal filter device Biomax-5K NMWL membrane. This material was applied at a flow rate of 0.5 ml/min to a Pharmacia HR 10/30 column that had been packed with Pharmacia Sephadex G-50 equilibrated in 10 mM sodium phosphate buffer (pH 7.0)/ 0.1 M NaCl. Fractions having the maximal phosphoramidon-sensitive protease activity were then pooled and centrifuged over a Millipore Ultrafree®-15 centrifugal filter device Biomax-50K NMWL membrane. Proteolytic activity analysis, *supra.*, indicated this material to have only phosphoramidon-sensitive protease activity. Pooling of the phosphoramidon-insensitive protease, the 58 kDa protein, was followed by concentrating in a Millipore Ultrafree®-15 centrifugal filter device Biomax-50K NMWL membrane and further separation on a Pharmacia Superdex-75 column. Fractions containing the protease were pooled.

Analysis of purified 58- and 41/38 kDa purified proteases revealed that, while both types of protease were completely inhibited with 1,10 phenanthroline, only the 41/38 kDa protease was inhibited with phosphoramidon. Further analysis of crude broth indicated that protease activity of day 1 W-14 broth has 23% of the total protease activity due to the 41/38 kDa protease, increasing to 44% in day three W-14 broth.

Standard SDS-PAGE analysis for examining protein purity and obtaining amino terminal sequence was performed using 4-20% gradient MiniPlus SeptraGels purchased from Integrated Separation Systems (Natick, MA). Proteins to be amino-terminal sequenced were blotted onto PVDF membrane following purification, *infra.*, (ProBlott™ Membranes; Applied Biosystems, Foster City, CA),

visualized with 0.1% amido black, excised, and sent to Cambridge Prochem; Cambridge, MA, for sequencing.

Deduced amino terminal sequence of the 58- (SEQ ID NO:45) and 41/38 kDa (SEQ ID NO:44) proteases from three day old W-14
5 broth were DV-GSEKANEKLK (SEQ ID NO: 45) and DSGDDDKVTNTDIHR (SEQ ID NO:44), respectively.

Sequencing of the 41/38 kDa protease revealed several amino termini, each one having an additional amino acid removed by proteolysis. Examination of the primary, secondary, tertiary and
10 quaternary sequences for the 38 and 41 kDa polypeptides allowed for deduction of the sequence shown above and revealed that these two proteases are homologous.

Example 11, Part A

15 Screening of *Photorhabdus* Genomic Library via use of Antibodies for Genes encoding TcbA Peptide

In parallel to the sequencing described above, suitable probing and sequencing was done based on the TcbA_{ii} peptide (SEQ
20 ID NO:1). This sequencing was performed by preparing bacterial culture broths and purifying the toxin as described in Examples 1 and 2 above.

Genomic DNA was isolated from the *Photorhabdus luminescens* strain W-14 grown in Grace's insect tissue culture medium. The
25 bacteria were grown in 5 ml of culture medium in a 250 ml Erlenmeyer flask at 28°C and 250 rpm for approximately 24 hours. Bacterial cells from 100 ml of culture medium were pelleted at 5000 x g for 10 minutes. The supernatant was discarded, and the cell pellets then were used for the genomic DNA isolation.

30 The genomic DNA was isolated using a modification of the CTAB method described in Section 2.4.3 of Ausubel (*supra.*). The section entitled "Large Scale CsCl prep of bacterial genomic DNA" was followed through step 6. At this point, an additional chloroform/isoamyl alcohol (24:1) extraction was performed
35 followed by a phenol/chloroform/isoamyl (25:24:1) extraction step and a final chloroform/isoamyl/alcohol (24:1) extraction. The DNA was precipitated by the addition of a 0.6 volume of isopropanol. The precipitated DNA was hooked and wound around the end of a bent glass rod, dipped briefly into 70% ethanol as a
40 final wash, and dissolved in 3 ml of TE buffer.

The DNA concentration, estimated by optical density at 280/260 nm, was approximately 2 mg/ml.

Using this genomic DNA, a library was prepared.

Approximately 50 ug of genomic DNA was partly digested with Sau3
5 Al. Then NaCl density gradient centrifugation was used to size
fractionate the partially digested DNA fragments. Fractions
containing DNA fragments with an average size of 12 kb, or
larger, as determined by agarose gel electrophoresis, were
ligated into the plasmid BluScript, Stratagene, La Jolla,
10 California, and transformed into an *E. coli* DH5 α or DHB10 strain.

Separately, purified aliquots of the protein were sent to
the biotechnology hybridoma center at the University of
Wisconsin, Madison for production of monoclonal antibodies to the
proteins. The material that was sent was the HPLC purified
15 fraction containing native bands 1 and 2 which had been denatured
at 65°C, and 20 μ g of which was injected into each of four mice.
Stable monoclonal antibody-producing hybridoma cell lines were
recovered after spleen cells from unimmunized mouse were fused
with a stable myeloma cell line. Monoclonal antibodies were
20 recovered from the hybridomas.

Separately, polyclonal antibodies were created by taking
native agarose gel purified band 1 (see Example 1) protein which
was then used to immunize a New Zealand white rabbit. The
protein was prepared by excising the band from the native agarose
25 gels, briefly heating the gel pieces to 65°C to melt the agarose,
and immediately emulsifying with adjuvant. Freund's complete
adjuvant was used for the primary immunizations and Freund's
incomplete was used for 3 additional injections at monthly
intervals. For each injection, approximately 0.2 ml of
30 emulsified band 1, containing 50 to 100 micrograms of protein,
was delivered by multiple subcutaneous injections into the back
of the rabbit. Serum was obtained 10 days after the final
injection and additional bleeds were performed at weekly
intervals for 3 weeks. The serum complement was inactivated by
35 heating to 56°C for 15 minutes and then stored at -20°C.

The monoclonal and polyclonal antibodies were then used to
screen the genomic library for the expression of antigens which
could be detected by the epitope. Positive clones were detected
on nitrocellulose filter colony lifts. An immunoblot analysis of
40 the positive clones was undertaken.

An analysis of the clones as defined by both immunoblot and Southern analysis resulted in the tentative identification of five classes of clones.

In the first class of clone was a gene encoding the peptide designated here as TcbA_{ii}. Full DNA sequence of this gene (TcbA) was obtained. It is set forth as SEQ ID NO:11. Confirmation that the sequence encodes the internal sequence of SEQ ID NO:1 is demonstrated by the presence of SEQ ID NO:1 at amino acid number 88 from the deduced amino acid sequence created by the open reading frame of SEQ ID NO:11. This can be confirmed by referring to SEQ ID NO:12, which is the deduced amino acid sequence created by SEQ ID NO:11.

The second class of toxin peptides contains the segments referred to above as TcaB_i, TcaB_{ii} and TcaC. Following the screening of the library with the polyclonal antisera, this second class of toxin genes was identified by several clones which produced different size proteins, all of which cross-reacted with the polyclonal antibody on an immunoblot and were also found to share DNA homology on a Southern Blot. Sequence comparison revealed that they belonged to the gene complex designated TcaB and TcaC above.

Three other classes of antibody toxin clones were also isolated in the polyclonal screen. These classes produced proteins that cross-react with a polyclonal antibody and also shared DNA homology with the classes as determined by Southern blotting. The classes have been designated Class III, Class IV and Class V. It was also possible to identify monoclonals that cross-reacted with Class I, II, III, and IV. This suggests that all have regions of high protein homology. Thus, it appears that the *P. luminescens* extracellular protein genes represent a family of genes which are evolutionarily related.

To further pursue the concept that there might be evolutionarily related variations in the toxin peptides contained within this organism, two approaches have been undertaken to examine other strains of *P. luminescens* for the presence of related proteins. This was done both by PCR amplification of genomic DNA and by immunoblot analysis using the polyclonal and monoclonal antibodies.

The results indicate that related proteins are produced by *P. luminescens* strains WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-11, WX-12, WX-15 and W-14.

5

Example 11, Part BSequence and analysis of Class III toxin clones - tcc

Further DNA sequencing was performed on plasmids isolated from Class III *E. coli* clones described in Example 11, Part A.

10 The nucleotide sequence was shown to be three closely linked open reading frames at this genomic locus. This locus was designated tcc with the three open reading frames designated tccA SEQ ID NO:56, tccB SEQ ID NO:58 and tccC SEQ ID NO:60 (Fig. 6B).

The deduced amino acid from the tccA open reading frame
15 indicates the gene encodes a protein of 105,459 Da. This protein was designated TccA. The first 12 amino acids of this protein match the N-terminal sequence obtained from a 108 kDa protein, SEQ ID NO:7, previously identified as part of the toxin complex.

The deduced amino acid from the tccB open reading frame
20 indicates this gene encodes a protein of 175,716 Da. This protein was designated TccB. The first 11 amino acids of this protein match the N-terminal sequence obtained from a protein with estimated molecular weight of 185 kDa, SEQ ID NO:8.

The deduced amino acid sequence of tccC indicated that this
25 open reading frame encodes a protein of 111,694 Da and the protein product was designated TccC.

Example 12Characterization of *Photorhabdus* Strains

30

In order to establish that the collection described herein was comprised of *Photorhabdus* strains, the strains herein were assessed in terms of recognized microbiological traits that are characteristic of *Photorhabdus* and which differentiate it from
35 other *Enterobacteriaceae* and *Xenorhabdus* spp. (Farmer, J.J. 1984. Bergey's Manual of Systemic Bacteriology, vol 1. pp. 510-511. (ed. Kreig N.R. and Holt, J.G.). Williams & Wilkins, Baltimore.; Akhurst and Boemare, 1988, Boemare et al., 1993). These characteristic traits are as follows: Gram's stain negative

rods, organism size of 0.5-2 μm in width and 2-10 μm in length, red/yellow colony pigmentation, presence of crystalline inclusion bodies, presence of catalase, inability to reduce nitrate, presence of bioluminescence, ability to take up dye from growth media, positive for protease production, growth-temperature range below 37°C, survival under anaerobic conditions and positively motile. (Table 18). Reference *Escherichia coli*, *Xenorhabdus* and *Photorhabdus* strains were included in all tests for comparison. The overall results are consistent with all strains being part of the family *Enterobacteriaceae* and the genus *Photorhabdus*.

A luminometer was used to establish the bioluminescence of each strain and provide a quantitative and relative measurement of light production. For measurement of relative light emitting units, the broths from each strain (cells and media) were measured at three time intervals after inoculation in liquid culture (6, 12, and 24 hr) and compared to background luminosity (uninoculated media and water). Prior to measuring light emission from the various broths, cell density was established by measuring light absorbance (560 nm) in a Gilford Systems (Oberlin, OH) spectrophotometer using a sipper cell. Appropriate dilutions were then made (to normalize optical density to 1.0 unit) before measuring luminosity. Aliquots of the diluted broths were then placed into cuvettes (300 μl each) and read in a Bio-Orbit 1251 Luminometer (Bio-Orbit Oy, Twiku, Finland). The integration period for each sample was 45 seconds. The samples were continuously mixed (spun in baffled cuvettes) while being read to provide oxygen availability. A positive test was determined as being ≥ 5 -fold background luminescence (~ 5 -10 units). In addition, colony luminosity was detected with photographic film overlays and visually, after adaptation in a darkroom. The Gram's staining characteristics of each strain were established with a commercial Gram's stain kit (BBL, Cockeysville, MD) used in conjunction with Gram's stain control slides (Fisher Scientific, Pittsburgh, PA). Microscopic evaluation was then performed using a Zeiss microscope (Carl Zeiss, Germany) 100X oil immersion objective lens (with 10X ocular and 2X body magnification). Microscopic examination of individual strains for organism size, cellular description and inclusion bodies (the latter after logarithmic growth) was

performed using wet mount slides (10X ocular, 2X body and 40X objective magnification) with oil immersion and phase contrast microscopy with a micrometer (Akhurst, R.J. and Boemare, N.E. 1990. Entomopathogenic Nematodes in Biological Control (ed. Gaugler, R. and Kaya, H.). pp. 75-90. CRC Press, Boca Raton, USA.; Baghdiguian S., Boyer-Giglio M.H., Thaler, J.O., Bonnot G., Boemare N. 1993. Biol. Cell 79, 177-185.). Colony pigmentation was observed after inoculation on Bacto nutrient agar, (Difco Laboratories, Detroit, MI) prepared as per label instructions.

10 Incubation occurred at 28°C and descriptions were produced after 5-7 days. To test for the presence of the enzyme catalase, a colony of the test organism was removed on a small plug from a nutrient agar plate and placed into the bottom of a glass test tube. One ml of a household hydrogen peroxide solution was gently

15 added down the side of the tube. A positive reaction was recorded when bubbles of gas (presumptive oxygen) appeared immediately or within 5 seconds. Controls of uninoculated nutrient agar and hydrogen peroxide solution were also examined. To test for nitrate reduction, each culture was inoculated into

20 10 ml of Bacto Nitrate Broth (Difco Laboratories, Detroit, MI). After 24 hours incubation at 28°C, nitrite production was tested by the addition of two drops of sulfanilic acid reagent and two drops of alpha-naphthylamine reagent (see Difco Manual, 10th edition, Difco Laboratories, Detroit, MI, 1984). The generation

25 of a distinct pink or red color indicates the formation of nitrite from nitrate. The ability of each strain to uptake dye from growth media was tested with Bacto MacConkey agar containing the dye neutral red; Bacto Tergitol-7 agar containing the dye bromothymol blue and Bacto EMB Agar containing the dye eosin-Y

30 (agars from Difco Laboratories, Detroit, MI, all prepared according to label instructions). After inoculation on these media, dye uptake was recorded after incubation at 28°C for 5 days. Growth on these latter media is characteristic for members of the family *Enterobacteriaceae*. Motility of each strain was

35 tested using a solution of Bacto Motility Test Medium (Difco Laboratories, Detroit, MI) prepared as per label instructions. A butt-stab inoculation was performed with each strain and motility was judged macroscopically by a diffuse zone of growth spreading from the line of inoculum. In many cases, motility was also

observed microscopically from liquid culture under wet mount slides. Biochemical nutrient evaluation for each strain was performed using BBL Enterotube II (Benton, Dickinson, Germany). Product instructions were followed with the exception that

5 incubation was carried out at 28°C for 5 days. Results were consistent with previously cited reports for *Photorhabdus*. The production of protease was tested by observing hydrolysis of gelatin using Bacto gelatin (Difco Laboratories, Detroit, MI) plates made as per label instructions. Cultures were inoculated

10 and the plates were incubated at 28°C for 5 days. To assess growth at different temperatures, agar plates (2% proteose peptone #3 with two percent Bacto-Agar (Difco, Detroit, MI) in deionized water) were streaked from a common source of inoculum. Plates were sealed with Nesco® film and incubated at 20, 28 and

15 37°C for up to three weeks. Plates showing no growth at 37°C showed no cell viability after transfer to a 28°C incubator for one week. Oxygen requirements for *Photorhabdus* strains were tested in the following manner. A butt-stab inoculation into fluid thioglycolate broth medium (Difco, Detroit, MI) was made.

20 The tubes were incubated at room temperature for one week and cultures were then examined for type and extent of growth. The indicator resazurin demonstrates the level of medium oxidation or the aerobiosis zone (Difco Manual, 10th edition, Difco Laboratories, Detroit, MI). Growth zone results obtained for the

25 *Photorhabdus* strains tested were consistent with those of a facultative anaerobic microorganism.

Table 18

Taxonomic Traits of *Photorhabdus* Strains

30

Strain	Traits Assessed*																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
W-14	<u>-</u>	+	+	<u>rd</u> <u>S</u>	+	-	+	+	+	<u>O</u>	+	+	+	+	+	+	-
WX-1	-	+	+	<u>rd</u> <u>S</u>	+	-	+	+	+	<u>O</u>	+	+	+	+	+	+	-
WX-2	-	+	+	<u>rd</u> <u>S</u>	+	-	+	+	+	<u>O</u>	+	+	+	+	+	+	-
WX-3	-	+	+	<u>rd</u> <u>S</u>	+	-	+	+	+	<u>YT</u>	+	+	+	+	+	+	-
WX-4	-	+	+	<u>rd</u> <u>S</u>	+	-	+	+	+	<u>YT</u>	+	+	+	+	+	+	-
WX-5	-	+	+	<u>rd</u> <u>S</u>	+	-	+	+	+	<u>LO</u>	+	+	+	+	+	+	-

WX-6	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	LY	+	+	+	+	+	+	-
WX-7	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	R	+	+	+	+	+	+	-
WX-8	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-
WX-9	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	YT	+	+	+	+	+	+	-
WX-10	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	Ro	+	+	+	+	+	+	-
WX-11	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	Ro	+	+	+	+	+	+	-
WX-12	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-
WX-14	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	LR	+	+	+	+	+	+	-
WX-15	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	LR	+	+	+	+	+	+	-
H9	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	LY	+	+	+	+	+	+	-
Hb	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	YT	+	+	+	+	+	+	-
Hm	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	TY	+	+	+	+	+	+	-
HP88	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	LY	+	+	+	+	+	+	-
NC-1	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-
W30	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	YT	+	+	+	+	+	+	-
WIR	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	RO	+	+	+	+	+	+	-
B2	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	R	+	+	+	+	+	+	-
43948	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-
43949	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-
43950	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-
43951	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-
43952	-	+	+	$\frac{rd}{S}$	+	-	+	+	+	O	+	+	+	+	+	+	-

* - A = Gram's stain, B=Crystalline inclusion bodies, C=Bioluminescence, D=Cell form, E=Motility, F=Nitrate reduction, G=Presence of catalase, H=Gelatin hydrolysis, I=Dye uptake, J=Pigmentation, K=Growth on EMB agar, L=Growth on MacConkey agar, M=Growth on Tergitol-7 agar, N=Facultative anaerobe, O=Growth at 20°C, P=Growth at 28°C, Q=Growth at 37°C, + - +/- = positive or negative for trait, rd=rod, S=sized within Genus descriptors, RO=red-orange, LR = light red, R= red, O= orange, Y= yellow, T= tan, LY= light yellow, YT= yellow tan, and LO= light orange.

10

Cellular fatty acid analysis is a recognized tool for bacterial characterization at the genus and species level (Tornabene, T.G. 1985. Lipid Analysis and the Relationship to

Chemotaxonomy in Methods in Microbiology, Vol 18, 209-244.;
Goodfellow, M. and O'Donnell, A.G. 1993. Roots of Bacterial
Systematics in Handbook of New Bacterial Systematics (ed.
Goodfellow, M. & O'Donnell, A.G.) pp. 3-54. London: Academic
5 Press Ltd.), these references are incorporated herein by
reference, and were used to confirm that our collection was
related at the genus level. Cultures were shipped to an
external, contract laboratory for fatty acid methyl ester
analysis (FAME) using a Microbial ID (MIDI, Newark, DE, USA)
10 Microbial Identification System (MIS). The MIS system consists of
a Hewlett Packard HP5890A gas chromatograph with a 25mm x 0.2mm
5% methylphenyl silicone fused silica capillary column. Hydrogen
is used as the carrier gas and a flame-ionization detector
functions in conjunction with an automatic sampler, integrator
15 and computer. The computer compares the sample fatty acid methyl
esters to a microbial fatty acid library and against a
calibration mix of known fatty acids. As selected by the
contract laboratory, strains were grown for 24 hours at 28°C on
trypticase soy agar prior to analysis. Extraction of samples was
20 performed by the contract lab as per standard FAME methodology.
There was no direct identification of the strains to any
luminescent bacterial group other than *Photorhabdus*. When the
cluster analysis was performed, which compares the fatty acid
profiles of a group of isolates, the strain fatty acid profiles
25 were related at the genus level.

The evolutionary diversity of the *Photorhabdus* strains in
our collection was measured by analysis of PCR (Polymerase Chain
Reaction) mediated genomic fingerprinting using genomic DNA from
each strain. This technique is based on families of repetitive
30 DNA sequences present throughout the genome of diverse bacterial
species (reviewed by Versalovic, J., Schneider, M., DE Bruijn,
F.J. and Lupski, J.R. 1994. *Methods Mol. Cell. Biol.*, 5, 25-40.).
Three of these, repetitive extragenic palindromic sequence (REP),
enterobacterial repetitive intergenic consensus (ERIC) and the
35 BOX element are thought to play an important role in the
organization of the bacterial genome. Genomic organization is
believed to be shaped by selection and the differential
dispersion of these elements within the genome of closely related
bacterial strains can be used to discriminate these strains (e.g.

Louws, F.J., Fulbright, D.W., Stephens, C.T. and DE Bruijn, F.J. 1994. Appl. Environ. Micro. 60, 2286-2295.). Rep-PCR utilizes oligonucleotide primers complementary to these repetitive sequences to amplify the variably sized DNA fragments lying
5 between them. The resulting products are separated by electrophoresis to establish the DNA "fingerprint" for each strain.

To isolate genomic DNA from our strains, cell pellets were resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to a
10 final volume of 10 ml and 12 ml of 5 M NaCl was then added. This mixture was centrifuged 20 min. at 15,000 x g. The resulting pellet was resuspended in 5.7 ml of TE and 300 µl of 10% SDS and 60 µl 20 mg/ml proteinase K (Gibco BRL Products, Grand Island, NY) were added. This mixture was incubated at 37 °C for 1 hr,
15 approximately 10 mg of lysozyme was then added and the mixture was incubated for an additional 45 min. One milliliter of 5M NaCl and 800 µl of CTAB/NaCl solution (10% w/v CTAB, 0.7 M NaCl) were then added and the mixture was incubated 10 min. at 65°C, gently agitated, then incubated and agitated for an additional 20 min.
20 to aid in clearing of the cellular material. An equal volume of chloroform/isoamyl alcohol solution (24:1, v/v) was added, mixed gently then centrifuged. Two extractions were then performed with an equal volume of phenol/chloroform/isoamyl alcohol (50:49:1). Genomic DNA was precipitated with 0.6 volume of isopropanol.
25 Precipitated DNA was removed with a glass rod, washed twice with 70% ethanol, dried and dissolved in 2 ml of STE (10 mM Tris-HCl pH8.0, 10 mM NaCl, 1 mM EDTA). The DNA was then quantitated by optical density at 260 nm. To perform rep-PCR analysis of *Photorhabdus* genomic DNA the following primers were used, REP1R-
30 I; 5'-IIIICGICGICATCIGGC-3' and REP2-I; 5'-ICGICTTATCIGGCCTAC-3'. PCR was performed using the following 25µl reaction: 7.75 µl H₂O, 2.5 µl 10X LA buffer (PanVera Corp., Madison, WI), 16 µl dNTP mix (2.5 mM each), 1 µl of each primer at 50 pM/µl, 1 µl DMSO, 1.5 µl genomic DNA (concentrations ranged from 0.075-0.480 µg/µl) and
35 0.25 µl TaKaRa EX Taq (PanVera Corp., Madison, WI). The PCR amplification was performed in a Perkin Elmer DNA Thermal Cycler (Norwalk, CT) using the following conditions: 95°C/7 min. then 35 cycles of; 94°C/1 min., 44°C/1 min., 65°C/8 min., followed by 15 min. at 65°C. After cycling, the 25 µl reaction was added to 5 µl

of 6X gel loading buffer (0.25% bromophenol blue, 40% w/v sucrose in H₂O). A 15x20cm 1%-agarose gel was then run in TBE buffer (0.09 M Tris-borate, 0.002 M EDTA) using 8 µl of each reaction. The gel was run for approximately 16 hours at 45v. Gels were then
5 stained in 20 µg/ml ethidium bromide for 1 hour and destained in TBE buffer for approximately 3 hours. Polaroid® photographs of the gels were then taken under UV illumination.

The presence or absence of bands at specific sizes for each strain was scored from the photographs and entered as a
10 similarity matrix in the numerical taxonomy software program, NTSYS-pc (Exeter Software, Setauket, NY). Controls of *E. coli* strain HB101 and *Xanthomonas oryzae* pv. *oryzae* assayed at the same time produced PCR "fingerprints" corresponding to published reports (Versalovic, J., Koeuth, T. and Lupski, J.R. 1991.
15 Nucleic Acids Res. 19, 6823-6831; Vera Cruz, C.M., Halda-Alija, L., Louws, F., Skinner, D.Z., George, M.L., Nelson, R.J., DE Bruijn, F.J., Rice, C. and Leach, J.E. 1995. Int. Rice Res. Notes, 20, 23-24.; Vera Cruz, C.M., Ardales, E.Y., Skinner, D.Z., Talag, J., Nelson, R.J., Louws, F.J., Leung, H., Mew, T.W. and
20 Leach, J.E. 1996. Phytopathology (in press, respectively). The data from *Photorhabdus* strains were then analyzed with a series of programs within NTSYS-pc; SIMQUAL (Similarity for Qualitative data) to generate a matrix of similarity coefficients (using the Jaccard coefficient) and SAHN (Sequential, Agglomerative,
25 Heirarchical and Nested) clustering [using the UPGMA (Unweighted Pair-Group Method with Arithmetic Averages) method] which groups related strains and can be expressed as a phenogram (Figure 5). The COPH (cophenetic values) and MXCOMP (matrix comparison) programs were used to generate a cophenetic value matrix and
30 compare the correlation between this and the original matrix upon which the clustering was based. A resulting normalized Mantel statistic (r) was generated which is a measure of the goodness of fit for a cluster analysis (r=0.8-0.9 represents a very good fit). In our case r = 0.919. Therefore, our collection is
35 comprised of a diverse group of easily distinguishable strains representative of the *Photorhabdus* genus.

Example 13
Insecticidal Utility of Toxin(s) Produced
by Various *Photorhabdus* Strains

5 Initial "seed" cultures of the various *Photorhabdus* strains were produced by inoculating 175 ml of 2% Proteose Peptone #3 (PP3) (Difco Laboratories, Detroit, MI) liquid media with a primary variant subclone in a 500 ml tribaffled flask with a Delong neck, covered with a Kaput. Inoculum for each seed culture
10 was derived from oil-overlay agar slant cultures or plate cultures. After inoculation, these flasks were incubated for 16 hrs at 28°C on a rotary shaker at 150 rpm. These seed cultures were then used as uniform inoculum sources for a given fermentation of each strain. Additionally, overlaying the post-
15 log seed culture with sterile mineral oil, adding a sterile magnetic stir bar for future resuspension and storing the culture in the dark, at room temperature provided long-term preservation of inoculum in a toxin-competent state. The production broths were inoculated by adding 1% of the actively growing seed culture
20 to fresh 2% PP3 media (e.g. 1.75 ml per 175 ml fresh media). Production of broths occurred in either 500 ml tribaffled flasks (see above), or 2800 ml baffled, convex bottom flasks (500 ml volume) covered by a silicon foam closure. Production flasks were incubated for 24-48 hrs under the above mentioned
25 conditions. Following incubation, the broths were dispensed into sterile 1 L polyethylene bottles, spun at 2600 x g for 1 hr at 10°C and decanted from the cell and debris pellet. The liquid broth was then vacuum filtered through Whatman GF/D (2.7 µm retention) and GF/B (1.0 µm retention) glass filters to remove
30 debris. Further broth clarification was achieved with a tangential flow microfiltration device (Pall Filtron, Northborough, MA) using a 0.5 µm open-channel filter. When necessary, additional clarification could be obtained by chilling the broth (to 4°C) and centrifuging for several hours at 2600 x
35 g. Following these procedures, the broth was filter sterilized using a 0.2 µm nitrocellulose membrane filter. Sterile broths were then used directly for biological assay, biochemical analysis or concentrated (up to 15-fold) using a 10,000 MW cut-off, M12 ultra-filtration device (Amicon, Beverly MA) or

centrifugal concentrators (Millipore, Bedford, MA and Pall Filtron, Northborough, MA) with a 10,000 MW pore size. In the case of centrifugal concentrators, the broth was spun at 2000 x g for approximately 2 hr. The 10,000 MW permeate was added to the corresponding retentate to achieve the desired concentration of components greater than 10,000 MW. Heat inactivation of processed broth samples was achieved by heating the samples at 100°C in a sand-filled heat block for 10 minutes.

The broth(s) and toxin complex(es) from different *Photobacterium* strains are useful for reducing populations of insects and were used in a method of inhibiting an insect population which comprises applying to a locus of the insect an effective insect inactivating amount of the active described. A demonstration of the breadth of insecticidal activity observed from broths of a selected group of *Photobacterium* strains fermented as described above is shown in Table 19. It is possible that additional insecticidal activities could be detected with these strains through increased concentration of the broth or by employing different fermentation methods. Consistent with the activity being associated with a protein, the insecticidal activity of all strains tested was heat labile (see above).

Culture broth(s) from diverse *Photobacterium* strains show differential insecticidal activity (mortality and/or growth inhibition, reduced adult emergence) against a number of insects. More specifically, the activity is seen against corn rootworm larvae and boll weevil larvae which are members of the insect order Coleoptera. Other members of the Coleoptera include wireworms, pollen beetles, flea beetles, seed beetles and Colorado potato beetle. Activity is also observed against aster leafhopper and corn plant hopper, which are members of the order Homoptera. Other members of the Homoptera include planthoppers, pear psylla, apple sucker, scale insects, whiteflies, spittle bugs as well as numerous host specific aphid species. The broths and purified toxin complex(es) are also active against tobacco budworm, tobacco hornworm and European corn borer which are members of the order Lepidoptera. Other typical members of this order are beet armyworm, cabbage looper, black cutworm, corn earworm, codling moth, clothes moth, Indian mealmoth, leaf rollers, cabbage worm, cotton bollworm, bagworm, Eastern tent

- caterpillar, sod webworm and fall armyworm. Activity is also seen against fruitfly and mosquito larvae which are members of the order *Diptera*. Other members of the order *Diptera* are, pea midge, carrot fly, cabbage root fly, turnip root fly, onion fly, crane fly and house fly and various mosquito species. Activity with broth(s) and toxin complex(es) is also seen against two-spotted spider mite which is a member of the order *Acarina* which includes strawberry spider mites, broad mites, citrus red mite, European red mite, pear rust mite and tomato russet mite.
- Activity against corn rootworm larvae was tested as follows. *Photorhabdus* culture broth(s) (0-15 fold concentrated, filter sterilized), 2% Proteose Peptone #3, purified toxin complex(es) [0.23 mg/ml] or 10 mM sodium phosphate buffer, pH 7.0 were applied directly to the surface (about 1.5 cm²) of artificial diet (Rose, R. I. and McCabe, J. M. (1973). *J. Econ. Entomol.* 66, (398-400) in 40 µl aliquots. Toxin complex was diluted in 10 mM sodium phosphate buffer, pH 7.0. The diet plates were allowed to air-dry in a sterile flow-hood and the wells were infested with single, neonate *Diabrotica undecimpunctata howardi* (Southern corn rootworm, SCR) hatched from surface sterilized eggs. The plates were sealed, placed in a humidified growth chamber and maintained at 27°C for the appropriate period (3-5 days). Mortality and larval weight determinations were then scored. Generally, 16 insects per treatment were used in all studies. Control mortality was generally less than 5%.
- Activity against boll weevil (*Anthonomus grandis*) was tested as follows. Concentrated (1-10 fold) *Photorhabdus* broths, control medium (2% Proteose Peptone #3), purified toxin complex(es) [0.23 mg/ml] or 10 mM sodium phosphate buffer, pH 7.0 were applied in 60 µl aliquots to the surface of 0.35 g of artificial diet (Stoneville Yellow lepidopteran diet) and allowed to dry. A single, 12-24 hr boll weevil larva was placed on the diet, and the wells were sealed and held at 25°C, 50% RH for 5 days. Mortality and larval weights were then assessed. Control mortality ranged between 0-13%.
- Activity against mosquito larvae was tested as follows. The assay was conducted in a 96-well microtiter plate. Each well contained 200 µl of aqueous solution (10-fold concentrated *Photorhabdus* culture broth(s), control medium (2% Proteose

Peptone #3), 10 mM sodium phosphate buffer, toxin complex(es) @ 0.23 mg/ml or H₂O) and approximately 20, 1-day old larvae (*Aedes aegypti*). There were 6 wells per treatment. The results were read at 3-4 days after infestation. Control mortality was
5 between 0-20%.

Activity against fruitflies was tested as follows. Purchased *Drosophila melanogaster* medium was prepared using 50% dry medium and a 50% liquid of either water, control medium (2% Proteose Peptone #3), 10-fold concentrated *Photorhabdus* culture
10 broth(s), purified toxin complex(es) [0.23 mg/ml] or 10 mM sodium phosphate buffer, pH 7.0. This was accomplished by placing 4.0 ml of dry medium in each of 3 rearing vials per treatment and adding 4.0 ml of the appropriate liquid. Ten late instar
15 *Drosophila melanogaster* maggots were then added to each 25 ml vial. The vials were held on a laboratory bench, at room temperature, under fluorescent ceiling lights. Pupal or adult counts were made after 15 days of exposure. Adult emergence as compared to water and control medium (0-16% reduction).

Activity against aster leafhopper adults (*Macrosteles*
20 *severini*) and corn planthopper nymphs (*Peregrinus maidis*) was tested with an ingestion assay designed to allow ingestion of the active without other external contact. The reservoir for the active/"food" solution is made by making 2 holes in the center of the bottom portion of a 35X10 mm Petri dish. A 2 inch Parafilm
25 M[®] square is placed across the top of the dish and secured with an "O" ring. A 1 oz. plastic cup is then infested with approximately 7 hoppers and the reservoir is placed on top of the cup, Parafilm down. The test solution is then added to the reservoir through the holes. In tests using 10-fold concentrated
30 *Photorhabdus* culture broth(s), the broth and control medium (2% Proteose Peptone #3) were dialyzed against 10 mM sodium phosphate buffer, pH 7.0 and sucrose (to 5%) was added to the resulting solution to reduce control mortality. Purified toxin complex(es) [0.23 mg/ml] or 10 mM sodium phosphate buffer, pH 7.0 was also
35 tested. Mortality is reported at day 3. The assay was held in an incubator at 28°C, 70% RH with a 16/8 photoperiod. The assays were graded for mortality at 72 hours. Control mortality was less than 6%.

Activity against lepidopteran larvae was tested as follows. Concentrated (10-fold) *Photorhabdus* culture broth(s), control medium (2% Proteose Peptone #3), purified toxin complex(es) [0.23 mg/ml] or 10 mM sodium phosphate buffer, pH 7.0 were applied
5 directly to the surface (~1.5 cm²) of standard artificial lepidopteran diet (Stoneville Yellow diet) in 40 ul aliquots. The diet plates were allowed to air-dry in a sterile flow-hood and each well was infested with a single, neonate larva. European corn borer (*Ostrinia nubilalis*) and tobacco hornworm (*Manduca*
10 *sexta*) eggs were obtained from commercial sources and hatched in-house, whereas tobacco budworm (*Heliothis virescens*) larvae were supplied internally. Following infestation with larvae, the diet plates were sealed, placed in a humidified growth chamber and maintained in the dark at 27°C for the appropriate period.
15 Mortality and weight determinations were scored at day 5. Generally, 16 insects per treatment were used in all studies. Control mortality generally ranged from 4-12.5% for control medium and was less than 10% for phosphate buffer.

Activity against two-spotted spider mite (*Tetranychus*
20 *urticae*) was determined as follows. Young squash plants were trimmed to a single cotyledon and sprayed to run-off with 10-fold concentrated broth(s), control medium (2% Proteose Peptone #3), purified toxin complex(es) [0.23 mg/ml] or 10 mM sodium phosphate buffer, pH 7.0. After drying, the plants were infested with a
25 mixed population of spider mites and held at lab temperature and humidity for 72 hr. Live mites were then counted to determine levels of control.

Table 19
Observed Insecticidal Spectrum of Broths From Different
Photorhabdus Strains

	<i>Photorhabdus</i> Strain	Sensitive* Insect Species
5	WX-1	3**, 4, 5, 6, 7, 8
	WX-2	2, 4
	WX-3	1, 4
	WX-4	1, 4
10	WX-5	4
	WX-6	4
	WX-7	3, 4, 5, 6, 7, 8
	WX-8	1, 2, 4
	WX-9	1, 2, 4
15	WX-10	4
	WX-11	1, 2, 4
	WX-12	2, 4, 5, 6, 7, 8
	WX-14	1, 2, 4
	WX-15	1, 2, 4
20	W30	3, 4, 5, 8
	NC-1	1, 2, 3, 4, 5, 6, 7, 8, 9
	WIR	2, 3, 5, 6, 7, 8
	HP88	1, 3, 4, 5, 7, 8
	Hb	3, 4, 5, 7, 8
25	Hm	1, 2, 3, 4, 5, 7, 8
	H9	1, 2, 3, 4, 5, 6, 7, 8
	W-14	1, 2, 3, 4, 5, 6, 7, 8, 10
	ATCC 43948	4
	ATCC 43949	4
30	ATCC 43950	4
	ATCC 43951	4
	ATCC 43952	4

* = $\geq 25\%$ mortality and/or growth inhibition vs. control

35 ** = 1; Tobacco budworm, 2; European corn borer, 3;
Tobacco hornworm, 4; Southern corn rootworm, 5;
Boll weevil, 6; Mosquito, 7; Fruit Fly, 8;
Aster Leafhopper, 9; Corn planthopper, 10;
Two-spotted spider mite.

Example 14Non W-14 Photorhabdus Strains:Purification, Characterization and Activity Spectrum5 Purification

The protocol, as follows, is similar to that developed for the purification of W-14 and was established based on purifying those fractions having the most activity against Southern corn root worm (SCR), as determined in bioassays (see Example 13).

10 Typically, 4-20 L of broth that had been filtered, as described in Example 13, were received and concentrated using an Amicon spiral ultra filtration cartridge Type SLY100 attached to an Amicon M-12 filtration device. The retentate contained native proteins consisting of molecular sizes greater than 100 kDa,

15 whereas the flow through material contained native proteins less than 100 kDa in size. The majority of the activity against SCR was contained in the 100 kDa retentate. The retentate was then continually diafiltered with 10 mM sodium phosphate (pH = 7.0) until the filtrate reached an $A_{280} < 0.100$. Unless otherwise

20 stated, all procedures from this point were performed in buffer as defined by 10 mM sodium phosphate (pH 7.0). The retentate was then concentrated to a final volume of approximately 0.20 L and filtered using a 0.45 mm Nalgene™ Filterware sterile filtration unit. The filtered material was loaded at 7.5 ml/min onto a

25 Pharmacia HR16/10 column which had been packed with PerSeptive Biosystem Poros® 50 HQ strong anion exchange matrix equilibrated in buffer using a PerSeptive Biosystem Sprint® HPLC system. After loading, the column was washed with buffer until an $A_{280} < 0.100$ was achieved. Proteins were then eluted from the column at

30 2.5 ml/min using buffer with 0.4 M NaCl for 20 min for a total volume of 50 ml. The column was then washed using buffer with 1.0 M NaCl at the same flow rate for an additional 20 min (final volume = 50 ml). Proteins eluted with 0.4 M and 1.0 M NaCl were placed in separate dialysis bags (Spectra/Por® Membrane MWCO:

35 2,000) and allowed to dialyze overnight at 4° C in 12 L buffer. The majority of the activity against SCR was contained in the 0.4 M fraction. The 0.4 M fraction was further purified by application of 20 ml to a Pharmacia XK 26/100 column that had been prepacked with Sepharose CL4B (Pharmacia) using a flow rate

of 0.75 ml/min. Fractions were pooled based on A280 peak profile and concentrated to a final volume of 0.75 ml using a Millipore Ultrafree®-15 centrifugal filter device Biomax-50K NMWL membrane. Protein concentrations were determined using a Biorad Protein Assay Kit with bovine gamma globulin as a standard.

Characterization

The native molecular weight of the SCR toxin complex was determined using a Pharmacia HR 16/50 that had been prepacked with Sepharose CL4B in buffer. The column was then calibrated using proteins of known molecular size thereby allowing for calculation of the toxin approximate native molecular size. As shown in Table 20, the molecular size of the toxin complex ranged from 777 kDa with strain Hb to 1,900 kDa with strain WX-14. The yield of toxin complex also varied, from strain WX-12 producing 0.8 mg/L to strain Hb, which produced 7.0 mg/L.

Proteins found in the toxin complex were examined for individual polypeptide size using SDS-PAGE analysis. Typically, 20 mg protein of the toxin complex from each strain was loaded onto a 2-15% polyacrylamide gel (Integrated Separation Systems) and electrophoresed at 20 mA in Biorad SDS-PAGE buffer. After completion of electrophoresis, the gels were stained overnight in Biorad Coomassie blue R-250 (0.2% in methanol: acetic acid: water; 40:10:40 v/v/v). Subsequently, gels were destained in methanol:acetic acid: water; 40:10:40 (v/v/v). The gels were then rinsed with water for 15 min and scanned using a Molecular Dynamics Personal Laser Densitometer®. Lanes were quantitated and molecular sizes were calculated as compared to Biorad high molecular weight standards, which ranged from 200-45 kDa.

Sizes of the individual polypeptides comprising the SCR toxin complex from each strain are listed in Table 21. The sizes of the individual polypeptides ranged from 230 kDa with strain WX-1 to a size of 16 kDa, as seen with strain WX-7. Every strain, with the exception of strain Hb, had polypeptides comprising the toxin complex that were in the 160-230 kDa range, the 100-160 kDa range, and the 50-80 kDa range. These data indicate that the toxin complex may vary in peptide composition and components from strain to strain, however, in all cases the

toxin attributes appears to consist of a large, oligomeric protein complex.

Table 20

5 Characterization of a Toxin Complex From
Non W-14 *Photorhabdus* Strains

Strain	Approx. Native Molecular Wt. ^a	Yield Active Fraction (mg/L) ^b
H9	972,000	1.8
Hb	777,000	7.0
Hm	1,400,000	1.1
HP88	813,000	2.5
NC1	1,092,000	3.3
WIR	979,000	1.0
WX-1	973,000	0.8
WX-2	951,000	2.2
WX-7	1,000,000	1.5
WX-12	898,000	0.4
WX-14	1,900,000	1.9
W-14	860,000	7.5
a Native molecular weight determined using a Pharmacia HR 16/50 column packed with Sepharose CL4B		
b Amount of toxin complex recovered from culture broth.		

Activity Spectrum

10 As shown in Table 21, the toxin complexes purified from strains Hm and H9 were tested for activity against a variety of insects, with the toxin complex from strain W-14 for comparison. The assays were performed as described in Example 13. The toxin complex from all three strains exhibited activity against tobacco
15 bud worm, European corn borer, Southern corn root worm, and aster leafhopper. Furthermore, the toxin complex from strains Hm and W-14 also exhibited activity against two-spotted spider mite. In addition, the toxin complex from W-14 exhibited activity against mosquito larvae. These data indicate that the toxin complex,
20 while having similarities in activities between certain orders of insects, can also exhibit differential activities against other orders of insects.

Table 21

The Approximate Sizes (in kDa) of Peptides in a Purified
Toxin Complex From Non W-14 *Photobacterium*

5

H9	Hb	Hm	HP	NC-1	WIR	WX-1	WX-2	WX-7	WX-12	WX-14	W-11
			88								
180	150	170	170	180	170	230	200	200	180	210	190
170	140	140	160	170	160	190	170	180	160	180	180
160	139	100	140	140	120	170	150	110	140	160	170
140	130	81	130	110	110	160	120	87	139	120	160
120	120	72	129	44	89	110	110	75	130	110	150
98	100	68	110	16	79	98	82	43	110	100	130
87	98	49	100		74	76	64	33	92	95	120
84	88	46	86		62	58	37	28	87	80	110
79	81	30	81		51	53	30	26	80	69	93
72	75	22	77		40	41		23	73	49	90
68	69	20	73		39	35		22	59	41	77
60	60	19	60		37	31		21	56	33	69
57	57		58		33	28		19	51		65
52	54		45		30	24		18	37		63
46	49		39		28	22		16	33		60
40	44		35		27				32		51
37	39				25				26		46
	37				23						40
	35										39
											29

Table 22

Observed Insecticidal Spectrum of a Purified Toxin Complex from
Photorhabdus Strains

5	<u><i>Photorhabdus</i></u> Strain	Sensitive* Insect Species
	Hm Toxin Complex	1**, 2, 3, 5, 6, 7, 8
	H9 Toxin Complex	1, 2, 3, 6, 7, 8
10	W-14 Toxin Complex	1, 2, 3, 4, 5, 6, 7, 8
	* = > 25% mortality or growth inhibition	
	* = > 25% mortality or growth inhibition	
15	** = 1; Tobacco bud worm, 2; European corn borer, 3; Southern corn root worm, 4; Mosquito, 5; Two-spotted spider mite, 6; Aster Leafhopper, 7; Fruit Fly, 8; Boll Weevil	

Example 15

Sub-Fractionation of *Photorhabdus* Protein Toxin Complex

20 The *Photorhabdus* protein toxin complex was isolated as described in Example 14. Next, about 10 mg toxin was applied to a MonoQ 5/5 column equilibrated with 20 mM Tris-HCl, pH 7.0 at a flow rate of 1ml/min. The column was washed with 20 mM Tris-HCl, pH 7.0 until the optical density at 280 nm returned to baseline absorbance. The proteins bound to the column were eluted with a linear gradient of 0 to 1.0 M NaCl in 20 mM Tris-HCl, pH 7.0 at 1 ml/min for 30 min. One ml fractions were collected and subjected to Southern corn rootworm (SCR) bioassay (see Example 13). Peaks of activity were determined by a series of dilutions of each fraction in SCR bioassays. Two activity peaks against SCR were observed and were named A (eluted at about 0.2-0.3 M NaCl) and B (eluted at 0.3-0.4 M NaCl). Activity peaks A and B were pooled separately and both peaks were further purified using a 3-step procedure described below.

35 Solid (NH₄)₂SO₄ was added to the above protein fraction to a final concentration of 1.7 M. Proteins were then applied to a phenyl-Superose 5/5 column equilibrated with 1.7 M (NH₄)₂SO₄ in 50 mM potassium phosphate buffer, pH 7 at 1 ml/min. Proteins bound to the column were eluted with a linear gradient of 1.7 M (NH₄)₂SO₄, 0% ethylene glycol, 50 mM potassium phosphate, pH 7.0 to 25% ethylene glycol, 25 mM potassium phosphate, pH 7.0 (no (NH₄)₂SO₄) at 0.5 ml/min. Fractions were dialyzed overnight

against 10 mM sodium phosphate buffer, pH 7.0. Activities in each fraction against SCR were determined by bioassay.

The fractions with the highest activity were pooled and applied to a MonoQ 5/5 column which was equilibrated with 20 mM Tris-HCl, pH 7.0 at 1 ml/min. The proteins bound to the column were eluted at 1 ml/min by a linear gradient of 0 to 1M NaCl in 20 mM Tris-HCl, pH 7.0.

For the final step of purification, the most active fractions above (determined by SCR bioassay) were pooled and subjected to a second phenyl-Superose 5/5/ column. Solid $(\text{NH}_4)_2\text{SO}_4$ was added to a final concentration of 1.7 M. The solution was then loaded onto the column equilibrated with 1.7 M $(\text{NH}_4)_2\text{SO}_4$ in 50 mM potassium phosphate buffer, pH 7 at 1ml/min. Proteins bound to the column were eluted with a linear gradient of 1.7 M $(\text{NH}_4)_2\text{SO}_4$, 50 mM potassium phosphate, pH 7.0 to 10 mM potassium phosphate, pH 7.0 at 0.5 ml/min. Fractions were dialyzed overnight against 10 mM sodium phosphate buffer, pH 7.0. Activities in each fraction against SCR were determined by bioassay.

The final purified protein by the above 3-step procedure from peak A was named toxin A and the final purified protein from peak B was named toxin B.

Characterization and Amino Acid Sequencing of Toxin A and Toxin B

In SDS-PAGE, both toxin A and toxin B contained two major (> 90% of total Commassie stained protein) peptides: 192 kDa (named A1 and B1, respectively) and 58 kDa (named A2 and B2, respectively). Both toxin A and toxin B revealed only one major band in native PAGE, indicating A1 and A2 were subunits of one protein complex, and B1 and B2 were subunits of one protein complex. Further, the native molecular weight of both toxin A and toxin B were determined to be 860 kDa by gel filtration chromatography. The relative molar concentrations of A1 to A2 was judged to be a 1 to 1 equivalence as determined by densitometric analysis of SDS-PAGE gels. Similarly, B1 and B2 peptides were present at the same molar concentration.

Toxin A and toxin B were electrophoresed in 10% SDS-PAGE and transblotted to PVDF membranes. Blots were sent for amino acid analysis and N-terminal amino acid sequencing at Harvard MicroChem and Cambridge ProChem, respectively. The N-terminal

amino sequence of B1 was determined to be identical to SEQ ID NO:1, the TcbA_{ii} region of the *tcbA* gene (SEQ ID NO:12, position 87 to 99). A unique N-terminal sequence was obtained for peptide B2 (SEQ ID NO:40). The N-terminal amino acid sequence of peptide
5 B2 was identical to the TcbA_{iii} region of the derived amino acid sequence for the *tcbA* gene (SEQ ID NO:12, position 1935 to 1945). Therefore, the B toxin contained predominantly two peptides, TcbA_{ii} and TcbA_{iii}, that were observed to be derived from the same gene product, TcbA.

10 The N-terminal sequence of A2 (SEQ ID NO:41) was unique in comparison to the TcbA_{iii} peptide and other peptides. The A2 peptide was denoted TcdA_{iii} (see Example 17). SEQ ID NO:6 was determined to be a mixture of amino acid sequences SEQ ID NO:40 and 41.

15 Peptides A1 and A2 were further subjected to internal amino acid sequencing. For internal amino acid sequencing, 10 µg of toxin A was electrophoresized in 10% SDS-PAGE and transblotted to PVDF membrane. After the blot was stained with amido black, peptides A1 and A2, denoted TcdA_{ii} and TcdA_{iii}, respectively,
20 were excised from the blot and sent to Harvard MicroChem and Cambridge ProChem. Peptides were subjected to trypsin digestion followed by HPLC chromatography to separate individual peptides. N-terminal amino acid analysis was performed on selected tryptic peptide fragments. Two internal amino acid sequences of peptide
25 A1 (TcdA_{ii}-PK71, SEQ ID NO:38 and TcdA_{ii}-PK44, SEQ ID NO:39) were found to have significant homologies with deduced amino acid sequences of the TcbA_{ii} region of the *tcbA* gene (SEQ ID NO:12). Similarly, the N-terminal sequence (SEQ ID NO:41) and two internal sequences of peptides A2 (TcdA_{iii}-PK57, SEQ ID NO:42 and
30 TcdA_{iii}-PK20, SEQ ID NO:43) also showed significant homology with deduced amino acid sequences of TcbA_{iii} region of the *tcbA* gene (SEQ ID NO:12).

In summary of above results, the toxin complex has at least two active protein toxin complexes against SCR; toxin A and toxin
35 B. Toxin A and toxin B are similar in their native and subunits molecular weight, however, their peptide compositions are different. Toxin A contained peptides TcdA_{ii} and TcdA_{iii} as the major peptides and the toxin B contains TcbA_{ii} and TcbA_{iii} as the major peptides.

Example 16Cleavage and Activation of TcbA Peptide

5

In the toxin B complex, peptide TcbA_{ii} and TcbA_{iii} originate from the single gene product TcbA (Example 15). The processing of TcbA peptide to TcbA_{ii} and TcbA_{iii} is presumably by the action of *Photorhabdus* protease(s), and most likely, the metalloproteases described in Example 10. In some cases, it was noted that when *Photorhabdus* W-14 broth was processed, TcbA peptide was present in toxin B complex as a major component, in addition to peptides TcbA_{ii} and TcbA_{iii}. Identical procedures, described for the purification of toxin B complex (Example 15), were used to enrich peptide TcbA from toxin complex fraction of W-14 broth. The final purified material was analyzed in a 4-20% gradient SDS-PAGE and major peptides were quantified by densitometry. It was determined that TcbA, TcbA_{ii} and TcbA_{iii} comprised 58%, 36%, and 6%, respectively, of total protein. The identities of these peptides were confirmed by their respective molecular sizes in SDS-PAGE and Western blot analysis using monospecific antibodies. The native molecular weight of this fraction was determined to be 860 kDa.

The cleavage of TcbA was evaluated by treating the above purified material with purified 38 kDa and 58 kDa W-14 *Photorhabdus* metalloproteases (Example 10), and Trypsin as a control enzyme (Sigma, MO). The standard reaction consisted 17.5 µg the above purified fraction, 1.5 unit protease, and 0.1 M Tris buffer, pH 8.0 in a total volume of 100 µl. For the control reaction, protease was omitted. The reaction mixtures were incubated at 37 °C for 90 min. At the end of the reaction, 20 µl was taken and boiled with SDS-PAGE sample buffer immediately for electrophoresis analysis in a 4-20% gradient SDS-PAGE. It was determined from SDS-PAGE that in both 38 kDa and 58 kDa protease treatments, the amount of peptides TcbA_{ii} and TcbA_{iii} increased about 3-fold while the amount of TcbA peptide decreased proportionally (Table 23). The relative reduction and augmentation of selected peptides was confirmed by Western blot analyses. Furthermore, gel filtration of the cleaved material revealed that the native molecular size of the complex remained the same. Upon trypsin treatment, peptides TcbA and TcbA_{ii} were

nonspecifically digested into small peptides. This indicated that 38 kDa and 58 kDa *Photorhabdus* proteases can specifically process peptide TcbA into peptides TcbA_{ii} and TcbA_{iii}. Protease treated and untreated control of the remaining 80 ul reaction mixture were serial diluted with 10 mM sodium phosphate buffer, pH 7.0 and analyzed by SCR bioassay. By comparing activity in several dilution, it was determined that the 38 kDa protease treatment increased SCR insecticidal activity approximately 3 to 4 fold. The growth inhibition of remaining insects in the protease treatment was also more severe than control (Table 23).

Table 23

Conversion and activation of peptide TcbA into peptides TcbA_{ii} and TcbA_{iii} by protease treatment.

	Control	38 kDa protease treatment
S0 (% of total protein)	58	18
S1 (% of total protein)	36	64
S9 (% of total protein)	6	18
LD50 (ug protein)	2.1	0.52
SCR Weight (mg/insect)*	0.2	0.1

*: an indication of growth inhibition by measuring the average weight of live insect after 5 days on diet in the assay.

25

Example 17Screening of the library for a gene encoding the TcdA_{ii} Peptide

The cloning and characterization of a gene encoding the TcdA_{ii} peptide, described as SEQ ID NO:17 (internal peptide TcdA_{ii}-PT111 N-terminal sequence) and SEQ ID NO:18 (internal peptide TcdA_{ii}-PT79 N-terminal sequence) was completed. Two pools of degenerate oligonucleotides, designed to encode the amino acid sequences of SEQ ID NO:17 (Table 24) and SEQ ID NO:18 (Table 25), and the reverse complements of those sequences, were synthesized as described in Example 8. The DNA sequence of the oligonucleotides is given below:

Table 24
Degenerate Oligonucleotide for SEQ ID NO:17

P2-PT111	1	2	3	4	5	6	7	8
Amino Acid	Ala	Phe	Asn	Ile	Asp	Asp	Val	Ser
Codons	5' GCN	TT(T/C)	AA(T/C)	AT(T/C/A)	GA(T/C)	GA(T/C)	GTN 3'	
P2.3.6.CB	5' GC(A/C/G/T)	TT(T/C)	AAT	ATT	GAT	GAT	GT 3'	
P2.3.5	5' GC(A/C/G/T)	TT(T/C)	AA(T/C)	AT(T/C/A)	GA(T/C)	GA(T/C)	GT 3'	
P2.3.5R	5' AC	(G/A)TC	(G/A)TC	(T/G/A)AT	(G/A)TT	(G/A)AA	(A/C/G/T)GC 3'	
P2.3.5RI	5' ACI	TCI	TCI	ATI	TTI	AAI	GC 3'	
P2.3R.CB	5' CAG	(A/G)CT	(A/C)AC	ATC	ATC	AAT	ATT	AAA 3'

Table 25
Degenerate Oligonucleotide for SEQ ID NO:18

P2-PT79	1	2	3	4	5	6	7	8	9	10	11	12	13
Amino Acid	Phe	Ile	Val	Tyr	Thr	Ser	Leu	Gly	Val	Asn	Pro	Asn	Asn
Codons*	5' TTY	ATH	GTN	TAY	ACN	6	6	GGN	GTN	AAV	CCN	AAV	AAV 3'
P2.79.2	5' TTY	ATY	GTK	TAT	ACY	TCI	YTR	GGY	GTK	AAT	CCR	AAT	AAT 3'
P2.79.3	5' TTT	ATT	GTK	TAT	ACY	AGY	YTR	GGY	GTK	AAT	CCR	AAT	AAT 3'
P2.79.R.1	5' ATT	ATT	YGG	ATT	MAC	RCC	VAR	RCT	RGT	ATA	MAC	AAT	AAA 3'
P2.79R.CB	5' ATT	ATT	YGG	ATT	MAC	ACC	CAG	RCT	GGT	ATA	MAC	AAT	AAA 3'

* According to IUPAC-IUB codes for nucleotides, Y = C or T, H = A, C or T,
N = A, C, G or T, K = G or T, R = A or G, and M = A or C

Polymerase Chain Reactions (PCR) were performed essentially as described in Example 8, using as forward primers P2.3.5.CB or P2.3.5, and as reverse primers P2.79.R.1 or P2.79R.CB, in all forward/reverse combinations, using *Photothabdus* W-14 genomic DNA as template. In another set of reactions, primers P2.79.2 or P2.79.3 were used as forward primers, and P2.3.5R, P2.3.5RI, and P2.3R.CB were used as reverse primers in all forward/reverse combinations. Only in the reactions containing P2.3.6.CB as the forward primers combined with P2.79.R.1 or P2.79R.CB as the reverse primers was a non-artifactual amplified product seen, of estimated size (mobility on agarose gels) of 2500 base pairs. The order of the primers used to obtain this amplification product indicates that the peptide fragment TcdA₁₁₁-PT111 lies amino-proximal to the peptide fragment TcdA₁₁₁-PT79.

The 2500 bp PCR products were ligated to the plasmid vector pCRTMII (Invitrogen, San Diego, CA) according to the supplier's instructions, and the DNA sequences across the ends of the insert fragments of two isolates (HS24 and HS27) were determined using the supplier's recommended primers and the sequencing methods described previously. The sequence of both isolates was the same. New primers were synthesized based on the determined sequence, and used to prime additional sequencing reactions to obtain a total of 2557 bases of the insert [SEQ ID NO:36]. Translation of the partial peptide encoded by SEQ ID No: 36 yields the 845 amino acid sequence disclosed as SEQ ID NO:37. Protein homology analysis of this portion of the TcdA₁₁₁ peptide fragment reveals substantial amino acid homology (68% similarity; 53% identity) to residues 542 to 1390 of protein TcbA [SEQ ID NO:12]. It is therefore apparent that the gene represented in part by SEQ ID NO:36 produces a protein of similar, but not identical, amino acid sequence as the TcbA protein, and which likely has similar, but not identical biological activity as the TcbA protein.

In yet another instance, a gene encoding the peptides TcdA₁₁₁-PK44 and the TcdA₁₁₁ 58 kDa N-terminal peptide, described as SEQ ID NO:9 (internal peptide TcdA₁₁₁-PK44 sequence), and SEQ ID NO:41 (TcdA₁₁₁ 58 kDa N-terminal peptide sequence) was isolated. Two pools of degenerate oligonucleotides, designed to encode the amino acid sequences described as SEQ ID NO:39 (Table 27) and SEQ

ID NO:41 (Table 26), and the reverse complements of those sequences, were synthesized as described in Example 3, and their DNA sequences.

Table 26
Degenerate Oligonucleotide for SEQ ID NO:41

Codon #	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Amino Acid	Leu	Arg	Ser	Ala	Asn	Thr	Leu	Thr	Asp	Leu	Phe	Leu	Pro	Gln
A2.1	5' YTR	CGY	AGY	GCI	ANT	ACY	YTR	ACY	GAT	YTR	TTT	YTR	OCR	CA 3'
A2.2				GCI	ANT	ACY	YTR	ACI	GAY	YTR	TTY	YTR	OCI	CA 3'
A2.3.R		5' TG	YCG	YAR	AAA	YAR	RIC	RGT	YAR	RGT	RUT	IGC	RCT	ROG 3'
A2.4.R				5' TG	ICG	CAG	AAA	CAG	RIC	IGT	CAG	IGT	ATT	IGC 3'

Table 27
Degenerate Oligonucleotide for SEQ ID NO:39

Amino Acid #	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Codon #	1	2	3	4	5	6	7	8	9
Amino Acid	Gly	Pro	Val	Glu	Ile	Asn	Thr	Ala	Ile
A1.44.1	5' GGY	CCR	GTK	GAA	ATT	AAT	ACC	GCI	AT 3'
A1.44.1R	5' ATI	GCG	GTA	TTA	ATT	TCM	ACY	GGR	CC 3'
A1.44.2	5' GGI	CCI	GTI	GAR	ATY	AAY	ACI	GCI	AT 3'
A1.44.2R	5' ATI	GCI	GTR	TTR	ATY	TCI	ACI	GCI	CC 3'

Polymerase Chain Reactions (PCR) were performed essentially as described in Example 8, using as forward primers A1.44.1 or A1.44.2, and reverse primers A2.3R or A2.4R, in all forward/reverse combinations, using *Photothabdus* W-14 genomic DNA as template. In another set of reactions, primers A2.1 or A2.2 were used as forward primers, and A1.44.1R, and A1.44.2R were used as reverse primers in all forward/reverse combinations. Only in the reactions containing A1.44.1 or A1.44.2 as the forward primers combined with A2.3R as the reverse primer was a non-artifactual amplified product seen, of estimated size (mobility on agarose gels) of 1400 base pairs. The order of the primers used to obtain this amplification product indicates that the peptide fragment TcdA_{iii}-PK44 lies amino-proximal to the 58 kDa peptide fragment of TcdA_{iii}.

The 1400 bp PCR products were ligated to the plasmid vector pCR[™]II according to the supplier's instructions. The DNA sequences across the ends of the insert fragments of four isolates were determined using primers similar in sequence to the supplier's recommended primers and using sequencing methods described previously. The nucleic acid sequence of all isolates differed as expected in the regions corresponding to the degenerate primer sequences, but the amino acid sequences deduced from these data were the same as the actual amino acid sequences for the peptides determined previously, (SEQ ID NOS:41 and 39).

Screening of the W-14 genomic cosmid library as described in Example 8 with a radiolabeled probe comprised of the DNA prepared above (SEQ ID NO:36) identified five hybridizing cosmid isolates, namely 17D9, 20B10, 21D2, 27B10, and 26D1. These cosmids were distinct from those previously identified with probes corresponding to the genes described as SEQ ID NO:11 or SEQ ID NO:25. Restriction enzyme analysis and DNA blot hybridizations identified three EcoR I fragments, of approximate sizes 3.7, 3.7, and 1.1 kbp, that span the region comprising the DNA of SEQ ID NO:36. Screening of the W-14 genomic cosmid library using as probe the radiolabeled 1.4 kbp DNA fragment prepared in this example identified the same five cosmids (17D9, 20B10, 21D2, 27B10, and 26D1). DNA blot hybridization to EcoR I-digested cosmid DNAs also showed hybridization to the same subset

of EcoR I fragments as seen with the 2.5 kbp TcdA_{iii} gene probe, indicating that both fragments are encoded on the genomic DNA.

DNA sequence determination of the cloned EcoR I fragments revealed an uninterrupted reading frame of 7551 base pairs (SEQ ID NO:46), encoding a 282.9 kDa protein of 2516 amino acids (SEQ ID NO:47). Analysis of the amino acid sequence of this protein revealed all expected internal fragments of peptides TcdA_{iii} (SEQ ID NOS:17, 18, 37, 38 and 39) and the TcdA_{iii} peptide N-terminus (SEQ ID NO:41) and all TcdA_{iii} internal peptides (SEQ ID NOS:42 and 43). The peptides isolated and identified as TcdA_{iii} and TcdA_{iii} are each products of the open reading frame, denoted tcdA, disclosed as SEQ ID NO:46. Further, SEQ ID NO:47 shows, starting at position 89, the sequence disclosed as SEQ ID NO:13, which is the N-terminal sequence of a peptide of size approximately 201 kDa, indicating that the initial protein produced from SEQ ID No: 46 is processed in a manner similar to that previously disclosed for SEQ ID NO:12. In addition, the protein is further cleaved to generate a product of size 209.2 kDa, encoded by SEQ ID NO:48 and disclosed as SEQ ID NO:49 (TcdA_{iii} peptide), and a product of size 63.6 kDa, encoded by SEQ ID NO:50 and disclosed as SEQ ID NO:51 (TcdA_{iii} peptide). Thus, it is thought that the insecticidal activity identified as toxin A (Example 15) derived from the products of SEQ ID NO:46, as exemplified by the full-length protein of 282.9 kDa disclosed as SEQ ID NO:47, is processed to produce the peptides disclosed as SEQ ID NOS:49 and 51. It is thought that the insecticidal activity identified as toxin B (Example 15) derives from the products of SEQ ID NO:11, as exemplified by the 280.6 kDa protein disclosed as SEQ ID NO:12. This protein is proteolytically processed to yield the 207.6 kDa peptide disclosed as SEQ ID NO:53, which is encoded by SEQ ID NO:52, and the 62.9 kDa peptide having N-terminal sequence disclosed as SEQ ID NO:40, and further disclosed as SEQ ID NO:55, which is encoded by SEQ ID NO:54.

Amino acid sequence comparisons between the proteins disclosed as SEQ ID NO:12 and SEQ ID NO:47 reveal that they have 69% similarity and 54% identity. This high degree of evolutionary relationship is not uniform throughout the entire amino acid sequence of these peptides, but is higher towards the carboxy-terminal end of the proteins, since the peptides

disclosed as SEQ ID NO:51 (derived from SEQ ID NO:47) and SEQ ID NO:55 (derived from SEQ ID NO:12) have 76% similarity and 64% identity.

5

Example 18

Control of European Cornborer-Induced Leaf Damage on Maize Plants
by Spray Application of *Photorhabdus* (Strain W-14) Broth

10 The ability of *Photorhabdus* toxin(s) to reduce plant damage
caused by insect larvae was demonstrated by measuring leaf damage
caused by European corn borer (*Ostrinia nubilalis*) infested onto
maize plants treated with *Photorhabdus* broth. Fermentation broth
from *Photorhabdus* strain W-14 was produced and concentrated
15 approximately 10-fold using ultrafiltration (10,000 MW pore-size)
as described in Example 13. The resulting concentrated broth was
then filter sterilized using 0.2 micron nitrocellulose membrane
filters. A similarly prepared sample of uninoculated 2% proteose
peptone #3 was used for control purposes. Maize plants (a
20 DowElanco proprietary inbred line) were grown from seed to
vegetative stage 7 or 8 in pots containing a soilless mixture in
a greenhouse (27°C day; 22°C night, about 50%RH, 14 hr day-
length, watered/fertilized as needed). The test plants were
arranged in a randomized complete block design (3 reps/treatment,
25 6 plants/treatment) in a greenhouse with temperature about 22°C
day; 18°C night, no artificial light and with partial shading,
about 50%RH and watered/fertilized as needed. Treatments
(uninoculated media and concentrated *Photorhabdus* broth) were
applied with a syringe sprayer, 2.0 mls applied from directly
30 (about 6 inches) over the whorl and 2.0 additional mls applied in
a circular motion from approximately one foot above the whorl.
In addition, one group of plants received no treatment. After
the treatments had dried (approximately 30 minutes), twelve
neonate European corn borer larvae (eggs obtained from commercial
35 sources and hatched in-house) were applied directly to the whorl.
After one week, the plants were scored for damage to the leaves
using a modified Guthrie Scale (Koziel, M. G., Beland, G. L.,
Bowman, C., Carozzi, N. B., Crenshaw, R., Crossland, L., Dawson,
J., Desai, N., Hill, M., Kadwell, S., Launis, K., Lewis, K.,
40 Maddox, D., McPherson, K., Meghji, M. Z., Merlin, E., Rhodes, R.,

Warren, G. W., Wright, M. and Evola, S. V. 1993).

Bio/Technology, 11, 194-195.) and the scores were compared statistically [T-test (LSD) $p < 0.05$ and Tukey's Studentized Range (HSD) Test $p < 0.1$]. The results are shown in Table 28. For reference, a score of 1 represents no damage, a score of 2 represents fine "window pane" damage on the unfurled leaf with no pinhole penetration and a score of 5 represents leaf penetration with elongated lesions and/or mid rib feeding evident on more than three leaves (lesions < 1 inch). These data indicate that broth or other protein containing fractions may confer protection against specific insect pests when delivered in a sprayable formulation or when the gene or derivative thereof, encoding the protein or part thereof, is delivered via a transgenic plant or microbe.

Table 28

Effect of *Photothabdus* Culture Broth on
European Corn Borer-Induced Leaf Damage on Maize

20	Treatment	Average Guthrie Score
	No Treatment	5.02 ^a
	Uninoculated medium	5.15 ^a
	<i>Photothabdus</i> Broth	2.24 ^b
25	Means with different letters are statistically different ($p < 0.05$ or $p < 0.1$).	

Example 19

Genetic Engineering of Genes for Expression in *E. coli*

30 Summary of constructions

A series of plasmids were constructed to express the *tcBA* gene of *Photothabdus* W-14 in *Escherichia coli*. A list of the plasmids is shown in Table 29. A brief description of each construction follows as well as a summary of the *E. coli* expression data obtained.

Table 29
Expression plasmids for the *tcba* gene.

Plasmid	Gene	Vector/Selection	Compartment
pDAB634	<i>tcba</i>	pBC/Chl	Intracellular
pAcGP67B/ <i>tcba</i>	<i>tcba</i>	pAcGP67B/Amp	Baculovirus, secreted
pDAB635	<i>tcba</i>	pET27b/Kan	Periplasm
pET15- <i>tcba</i>	<i>tcba</i>	pET15- <i>tcba</i>	Intracellular

Abbreviations: Kan=kanamycin, Chl=chloramphenicol, Amp=ampicillin

5

Construction of pDAB634

In Example 9, a large *EcoR* I fragment which hybridizes to the *TcbA_{ii}* probe is described. This fragment was subcloned into pBC (Stratagene, La Jolla CA). Sequence analysis indicates that this fragment is 8816 base pairs. The fragment encodes the *tcba* gene with the initiating ATG at position 571 and the terminating TAA at position 8086. The fragment therefore carries 570 base pairs of *Photorhabdus* DNA upstream of the ATG and 730 base pairs downstream of the TAA.

15

Construction of Plasmid pAcGP67B/*tcba*

The *tcba* gene was PCR amplified using the following primers: 5' primer (SlAc51) 5' TTT AAA CCA TGG GAA ACT CAT TAT CAA GCA CTA TC 3' and 3' primer (SlAc31) 5' TTT AAA GCG GCC GCT TAA CGG ATG GTA TAA CGA ATA TG 3'. PCR was performed using a TaKaRa LA PCR kit from PanVera (Madison, Wisconsin) in the following reaction: 57.5 ml water, 10 ml 10X LA buffer, 16 ml dNTPs (2.5 mM each stock solution), 20 ml each primer at 10 pmoles/ml, 300 ng of the plasmid pDAB634 containing the W-14 *tcba* gene and one ml of TaKaRa LA Taq polymerase. The cycling conditions were 98°C/20 sec, 68°C/5 min, 72°C/10 min for 30 cycles. A PCR product of the expected about 7526bp was isolated in a 0.8% agarose gel in TBE (100 mM Tris, 90 mM boric acid, 1 mM EDTA) buffer and purified using a Qiaex II kit from Qiagen (Chatsworth, California). The purified *tcba* gene was digested with *Nco* I and *Not* I and ligated into the baculovirus transfer vector pAcGP67B (PharMingen (San Diego, California)) and transformed into DH5 α *E. coli*. The *tcba* gene was then cut from pAcGP67B and transferred to pET27b to create plasmid pDAB635. A missense mutation in the *tcba* gene was repaired in pDAB635.

The repaired *tcbA* gene contains two changes from the sequence shown in Sequence ID NO:11; an A>G at 212 changing an asparagine 71 to serine 71 and a G>A at 229 changing an alanine 77 to threonine 77. These changes are both upstream of the proposed TcbA_{ij} N-terminus.

Construction of pET15-*tcbA*

The *tcbA* coding region of pDAB635 was transferred to vector pET15b. This was accomplished using shotgun ligations, the DNAs were cut with restriction enzymes Nco I and Xho I. The resulting recombinant is called pET15-*tcbA*.

Expression of TcbA in *E. coli* from plasmid pET15-*tcbA*

Expression of *tcbA* in *E. coli* was obtained by modification of the methods previously described by Studier et al. (Studier, F.W., Rosenberg, A., Dunn, J., and Dubendorff, J., (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, 185: 60-89.). Competent *E. coli* cells strain BL21(DE3) were transformed with plasmid pET15-*tcbA* and plated on LB agar containing 100 µg/ml ampicillin and 40 mM glucose. The transformed cells were plated to a density of several hundred isolated colonies/plate. Following overnight incubation at 37°C the cells were scraped from the plates and suspended in LB broth containing 100 µg /ml ampicillin. Typical culture volumes were from 200-500 ml. At time zero, culture densities (OD600) were from 0.05-0.15 depending on the experiment. Cultures were shaken at one of three temperatures (22°C, 30°C or 37°C) until a density of 0.15-0.5 was obtained at which time they were induced with 1 mM isopropylthio-β-galactoside (IPTG). Cultures were incubated at the designated temperature for 4-5 hours and then were transferred to 4°C until processing (12-72 hours).

Purification and characterization of TcbA expressed in *E. coli* from Plasmid pET15-*tcbA*.

E. coli cultures expressing TcbA peptides were processed as follows. Cells were harvested by centrifugation at 17,000 x G and the media was decanted and saved in a separate container.

The media was concentrated about 8x using the M12 (Amicon, Beverly MA) filtration system and a 100 kD molecular mass cut-off filter. The concentrated media was loaded onto an anion exchange

column and the bound proteins were eluted with 1.0 M NaCl. The 1.0 M NaCl elution peak was found to cause mortality against Southern corn rootworm (SCR) larvae (Table 30). The 1.0 M NaCl fraction was dialyzed against 10 mM sodium phosphate buffer pH 7.0, concentrated, and subjected to gel filtration on Sepharose CL-4B (Pharmacia, Piscataway, New Jersey). The region of the CL-4B elution profile corresponding to calculated molecular weight (about 900 kDa) as the native W-14 toxin complex was collected, concentrated and bioassayed against larvae. The collected 900 kDa fraction was found to have insecticidal activity (see Table 30 below), with symptomology similar to that caused by native W-14 toxin complex. This fraction was subjected to Proteinase K and heat treatment, the activity in both cases was either eliminated or reduced, providing evidence that the activity is proteinaceous in nature. In addition, the active fraction tested immunologically positive for the TcbA and TcbA_{iii} peptides in immunoblot analysis when tested with an anti-TcbA_{iii} monoclonal antibody (Table 30).

Table 30
Results of Immunoblot and SCR Bioassays.

Fraction	SCR Activity		Immunoblot Peptides Detected	Native Size [CL-4B Estimated Size]
	% Mortality	% Growth Inhibit.		
TcbA Media 1.0 M Ion Exchange	+++	+++	TcbA	
TcbA Media CL-4B	+++	+++	TcbA, TcbA _{iii}	~900 kDa
TcbA Media CL-4B + Proteinase K	++	+++	NT	
TcbA Media CL-4B + heat treatment	-	-	NT	
TcbA Cell Sup CL-4B	-	+++	NT	~900 kD

PK = Proteinase K treatment 2 hours; Heat treatment = 100°C for 10 minutes; ND = None Detected; NT = Not Tested. Scoring system for mortality and growth inhibition as compared to control samples; 5-24%="+", 25-49%="++", 50-100%="+++".

The cell pellet was resuspended in 10 mM sodium phosphate buffer, pH=7.0, and lysed by passage through a Bio-Neb™ cell nebulizer (Glas-Col Inc., Terra Haute, IN). The pellets were

treated with DNase to remove DNA and centrifuged at $17,000 \times g$ to separate the cell pellet from the cell supernatant. The supernatant fraction was decanted and filtered through a 0.2 micron filter to remove large particles and subjected to anion exchange chromatography. Bound proteins were eluted with 1.0 M NaCl, dialyzed and concentrated using Biomax™ (Millipore Corp, Bedford, MA) concentrators with a molecular mass cut-off of 50,000 Daltons. The concentrated fraction was subjected to gel filtration chromatography using Sepharose CL-4B beaded matrix. Bioassay data for material prepared in this way is shown in Table 30 and is denoted as "TcbA Cell Sup".

In yet another method to handle large amounts of material, the cell pellets were re-suspended in 10 mM sodium phosphate buffer, pH = 7.0 and thoroughly homogenized by using a Kontes Glass Company (Vineland, NJ) 40 ml tissue grinder. The cellular debris was pelleted by centrifugation at $25,000 \times g$ and the cell supernatant was decanted, passed through a 0.2 micron filter and subjected to anion exchange chromatography using a Pharmacia 10/10 column packed with Poros HQ 50 beads. The bound proteins were eluted by performing a NaCl gradient of 0.0 to 1.0 M. Fractions containing the TcbA protein were combined and concentrated using a 50 kDa concentrator and subjected to gel filtration chromatography using Pharmacia CL-4B beaded matrix. The fractions containing TcbA oligomer, molecular mass of approximately 900 kDa, were collected and subjected to anion exchange chromatography using a Pharmacia Mono Q 10/10 column equilibrated with 20 mM Tris buffer pH = 7.3. A gradient of 0.0 to 1.0 M NaCl was used to elute recombinant TcbA protein. Recombinant TcbA eluted from the column at a salt concentration of approximately 0.3-0.4 M NaCl, the same molarity at which native TcbA oligomer is eluted from the Mono Q 10/10 column. The recombinant TcbA fraction was found to cause SCR mortality in bioassay experiments similar to those in Table 30.

35

SEQUENCE LISTING

- 5 (i) GENERAL INFORMATION:
- (i) APPLICANT: Ensign, Jerald C
Bowen, David J
Petell, James
Fatig, Raymond
10 Schoonover, Sue
ffrench-Constant, Richard
Orr, Gregory L
Merlo, Donald J
Roberts, Jean L
15 Rocheleau, Thomas A
Blackburn, Michael B
Hey, Timothy D
Strickland, James A
- 20 (ii) TITLE OF INVENTION: Insecticidal Protein Toxins From
Photorhabdus
- (iii) NUMBER OF SEQUENCES: 61
- 25 (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Quarles & Brady
(B) STREET: 1 South Pinckney Street
(C) CITY: Madison
(D) STATE: WI
30 (E) COUNTRY: US
(F) ZIP: 53703
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
35 (B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
40 (A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
45 (A) APPLICATION NUMBER: US 08/063,615
(B) FILING DATE: 18-MAY-1993
- (vii) PRIOR APPLICATION DATA:
50 (A) APPLICATION NUMBER: US 08/395,497
(B) FILING DATE: 28-FEB-1995
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/007,255
55 (B) FILING DATE: 06-NOV-1995
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/608,423
(B) FILING DATE: 28-FEB-1996

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/705,484
- (B) FILING DATE: 28-AUG-1996

5

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Seay, Nicholas J
- (B) REGISTRATION NUMBER: 27386
- (C) REFERENCE/DOCKET NUMBER: 960296.93804

10

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 608-251-5000
- (B) TELEFAX: 608-251-9166

15

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: protein

25

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

30

Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn
1 5 10

35

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: protein

45

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

50

Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Trp
1 5 10

(2) INFORMATION FOR SEQ ID NO:3:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

60

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp Ala
1 5 10 15

10

Leu Val Ala

15 (2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

20

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

25

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

30

Ala Ser Pro Leu Ser Thr Ser Glu Leu Thr Ser Lys Leu Asn
1 5 10

35 (2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

40

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

45

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

50

Ala Gly Asp Thr Ala Asn Ile Gly Asp
1 5

(2) INFORMATION FOR SEQ ID NO:6:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

60

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Leu Gly Gly Ala Ala Thr Leu Leu Asp Leu Leu Leu Pro Gln Ile
1 5 10 15

10

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu
1 5 10

30 (2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

40 (v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

45 Met Asn Leu Ala Ser Pro Leu Ile Ser
1 5

50 (2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

60 (v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ile Asn Leu Asp Ile Asn Glu Gln Asn Lys Ile Met Val Val Ser
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ala Ala Lys Asp Val Lys Phe Gly Ser Asp Ala Arg Val Lys Met Leu
 1 5 10 15
 Arg Gly Val Asn
 20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7515 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 1..7515

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATG CAA AAC TCA TTA TCA AGC ACT ATC GAT ACT ATT TGT CAG AAA CTG 48
 Met Gln Asn Ser Leu Ser Ser Thr Ile Asp Thr Ile Cys Gln Lys Leu
 1 5 10 15
 CAA TTA ACT TGT CCG GCG GAA ATT GCT TTG TAT CCC TTT GAT ACT TTC 96
 Gln Leu Thr Cys Pro Ala Glu Ile Ala Leu Tyr Pro Phe Asp Thr Phe
 20 25 30
 CGG GAA AAA ACT CGG GGA ATG GTT AAT TGG GGG GAA GCA AAA CGG ATT 144
 Arg Glu Lys Thr Arg Gly Met Val Asn Trp Gly Glu Ala Lys Arg Ile
 35 40 45
 TAT GAA ATT GCA CAA GCG GAA CAG GAT AGA AAC CTA CTT CAT GAA AAA 192
 Tyr Glu Ile Ala Gln Ala Glu Gln Asp Arg Asn Leu Leu His Glu Lys
 50 55 60
 CGT ATT TTT GCC TAT GCT AAT CCG CTG CTG AAA AAC GCT GTT CGG TTC 240
 Arg Ile Phe Ala Tyr Ala Asn Pro Leu Leu Lys Asn Ala Val Arg Leu

	65				70					75					80		
5	GGT Gly	ACC Thr	CGG Arg	CAA Gln	ATG Met 85	TTG Leu	GGT Gly	TTT Phe	ATA Ile	CAA Gln 90	GGT Gly	TAT Tyr	AGT Ser	GAT Asp	CTG Leu 95	TTT Phe	288
10	GGT Gly	AAT Asn	CGT Arg	GCT Ala 100	GAT Asp	AAC Asn	TAT Tyr	GCC Ala 105	GCG Ala	CCG Pro	GGC Gly	TCG Ser	GTT Val	GCA Ala 110	TCG Ser	ATG Met	336
15	TTG Phe	TCA Ser	CCG Pro 115	GCG Ala	GCT Ala	TAT Tyr	TTG Leu	ACG Thr 120	GAA Glu	TTG Leu	TAC Tyr	CGT Arg	GAA Glu 125	GCC Ala	AAA Lys	AAC Asn	384
20	TTG Leu	CAT His 130	GAC Asp	AGC Ser	AGC Ser	TCA Ser	ATT Ile 135	TAT Tyr	TAC Tyr	CTA Leu	GAT Asp	AAA Lys 140	CGT Arg	CGC Arg	CCG Pro	GAT Asp	432
25	TTA Leu	GCA Ala	AGC Ser	TTA Leu	ATG Met	CTC Leu 150	AGC Ser	CAG Gln	AAA Lys	AAT Asn	ATG Met 155	GAT Asp	GAG Glu	GAA Glu	ATT Ile	TCA Ser 160	480
30	ACG Thr	CTG Leu	GCT Ala	CTC Leu	TCT Ser 165	AAT Asn	GAA Glu	TTG Leu	TGC Cys	CTT Leu 170	GCC Ala	GGG Gly	ATC Ile	GAA Glu 175	ACA Thr	AAA Lys	528
35	ACA Thr	GGA Gly	AAA Lys	TCA Ser 180	CAA Gln	GAT Asp	GAA Glu	GTG Val	ATG Met 185	GAT Asp	ATG Met	TTG Leu	TCA Ser	ACT Thr 190	TAT Tyr	CGT Arg	576
40	TTA Leu	AGT Ser	GGA Gly 195	GAG Glu	ACA Thr	CCT Pro	TAT Tyr	CAT His 200	CAC His	GCT Ala	TAT Tyr	GAA Glu	ACT Thr 205	GTT Val	CGT Arg	GAA Glu	624
45	ATC Ile	GTT Val 210	CAT His	GAA Glu	CGT Arg	GAT Asp	CCA Pro 215	GGA Gly	TTT Phe	CGT Arg	CAT His	TTG Leu 220	TCA Ser	CAG Gln	GCA Ala	CCC Pro	672
50	ATT Ile	GTT Val 225	GCT Ala	GCT Ala	AAG Lys	CTC Leu 230	GAT Asp	CCT Pro	GTG Val	ACT Thr	TTG Leu 235	TTG Leu	GGT Gly	ATT Ile	AGC Ser	TCC Ser 240	720
55	CAT His	ATT Ile	TCG Ser	CCA Pro	GAA Glu 245	CTG Leu	TAT Tyr	AAC Asn	TTG Leu	CTG Leu 250	ATT Ile	GAG Glu	GAG Glu	ATC Ile	CCG Pro 255	GAA Glu	768
60	AAA Lys	GAT Asp	GAA Glu	GCC Ala 260	GCG Ala	CTT Leu	GAT Asp	ACG Thr	CTT Leu 265	TAT Tyr	AAA Lys	ACA Thr	AAC Asn	TTT Phe 270	GGC Gly	GAT Asp	816
65	ATT Ile	ACT Thr	ACT Thr	GCT Ala 275	CAG Gln	TTA Leu	ATG Met	TCC Ser 280	CCA Pro	AGT Ser	TAT Tyr	CTG Leu	GCC Ala 285	CGG Arg	TAT Tyr	TAT Tyr	864
70	GGC Gly	GTC Val 290	TCA Ser	CCG Pro	GAA Glu	GAT Asp	ATT Ile 295	GCC Ala	TAC Tyr	GTG Val	ACG Thr	ACT Thr 300	TCA Ser	TTA Leu	TCA Ser	CAT His	912
75	GTT Val	GGA Gly	TAT Tyr	AGC Ser	AGT Ser	GAT Asp 310	ATT Ile	CTG Leu	GTT Val	ATT Ile	CCG Pro 315	TTG Leu	GTC Val	GAT Asp	GGT Gly	GTG Val 320	960
80	GGT Gly	AAG Lys	ATG Met	GAA Glu	GTA Val 325	GTT Val	CGT Arg	GTT Val	ACC Thr	CGA Arg 330	ACA Thr	CCA Pro	TCG Ser	GAT Asp	AAT Asn 335	TAT Tyr	1008

	ACC	AGT	CAG	ACG	AAT	TAT	ATT	GAG	CTG	TAT	CCA	CAG	GGT	GGC	GAC	AAT	1056
	Thr	Ser	Gln	Thr	Asn	Tyr	Ile	Glu	Leu	Tyr	Pro	Gln	Gly	Gly	Asp	Asn	
				340					345					350			
5	TAT	TTG	ATC	AAA	TAC	AAT	CTA	AGC	AAT	AGT	TTT	GGT	TTG	GAT	GAT	TTT	1104
	Tyr	Leu	Ile	Lys	Tyr	Asn	Leu	Ser	Asn	Ser	Phe	Gly	Leu	Asp	Asp	Phe	
			355					360					365				
10	TAT	CTG	CAA	TAT	AAA	GAT	GGT	TCC	GCT	GAT	TGG	ACT	GAG	ATT	GCC	CAT	1152
	Tyr	Leu	Gln	Tyr	Lys	Asp	Gly	Ser	Ala	Asp	Trp	Thr	Glu	Ile	Ala	His	
			370				375					380					
15	AAT	CCC	TAT	CCT	GAT	ATG	GTC	ATA	AAT	CAA	AAG	TAT	GAA	TCA	CAG	GCG	1200
	Asn	Pro	Tyr	Pro	Asp	Met	Val	Ile	Asn	Gln	Lys	Tyr	Glu	Ser	Gln	Ala	
	385					390					395					400	
20	ACA	ATC	AAA	CGT	AGT	GAC	TCT	GAC	AAT	ATA	CTC	AGT	ATA	GGG	TTA	CAA	1248
	Thr	Ile	Lys	Arg	Ser	Asp	Ser	Asp	Asn	Ile	Leu	Ser	Ile	Gly	Leu	Gln	
				405					410						415		
25	AGA	TGG	CAT	AGC	GGT	AGT	TAT	AAT	TTT	GCC	GCC	GCC	AAT	TTT	AAA	ATT	1296
	Arg	Trp	His	Ser	Gly	Ser	Tyr	Asn	Phe	Ala	Ala	Ala	Asn	Phe	Lys	Ile	
				420					425					430			
30	GAC	CAA	TAC	TCC	CCG	AAA	GCT	TTC	CTG	CTT	AAA	ATG	AAT	AAG	GCT	ATT	1344
	Asp	Gln	Tyr	Ser	Pro	Lys	Ala	Phe	Leu	Leu	Lys	Met	Asn	Lys	Ala	Ile	
			435				440						445				
35	CGG	TTG	CTC	AAA	GCT	ACC	GGC	CTC	TCT	TTT	GCT	ACG	TTG	GAG	CGT	ATT	1392
	Arg	Leu	Leu	Lys	Ala	Thr	Gly	Leu	Ser	Phe	Ala	Thr	Leu	Glu	Arg	Ile	
		450					455					460					
40	GTT	GAT	AGT	GTT	AAT	AGC	ACC	AAA	TCC	ATC	ACG	GTT	GAG	GTA	TTA	AAC	1440
	Val	Asp	Ser	Val	Asn	Ser	Thr	Lys	Ser	Ile	Thr	Val	Glu	Val	Leu	Asn	
	465				470						475				480		
45	AAG	GTT	TAT	CGG	GTA	AAA	TTC	TAT	ATT	GAT	CGT	TAT	GGC	ATC	AGT	GAA	1488
	Lys	Val	Tyr	Arg	Val	Lys	Phe	Tyr	Ile	Asp	Arg	Tyr	Gly	Ile	Ser	Glu	
				485					490						495		
50	GAG	ACA	GCC	GCT	ATT	TTG	GCT	AAT	ATT	AAT	ATC	TCT	CAG	CAA	GCT	GTT	1536
	Glu	Thr	Ala	Ala	Ile	Leu	Ala	Asn	Ile	Asn	Ile	Ser	Gln	Gln	Ala	Val	
				500				505						510			
55	GGC	AAT	CAG	CTT	AGC	CAG	TTT	GAG	CAA	CTA	TTT	AAT	CAC	CCG	CCG	CTC	1584
	Gly	Asn	Gln	Leu	Ser	Gln	Phe	Glu	Gln	Leu	Phe	Asn	His	Pro	Pro	Leu	
			515				520						525				
60	AAT	GGT	ATT	CGC	TAT	GAA	ATC	AGT	GAG	GAC	AAC	TCC	AAA	CAT	CTT	CCT	1632
	Asn	Gly	Ile	Arg	Tyr	Glu	Ile	Ser	Glu	Asp	Asn	Ser	Lys	His	Leu	Pro	
		530				535						540					
65	AAT	CCT	GAT	CTG	AAC	CTT	AAA	CCA	GAC	AGT	ACC	GGT	GAT	GAT	CAA	CGC	1580
	Asn	Pro	Asp	Leu	Asn	Leu	Lys	Pro	Asp	Ser	Thr	Gly	Asp	Asp	Gln	Arg	
	545					550					555					560	
70	AAG	GCG	GTT	TTA	AAA	CGC	GCG	TTT	CAG	GTT	AAC	GCC	AGT	GAG	TTG	TAT	1728
	Lys	Ala	Val	Leu	Lys	Arg	Ala	Phe	Gln	Val	Asn	Ala	Ser	Glu	Leu	Tyr	
				565					570						575		
75	CAG	ATG	TTA	TTG	ATC	ACT	GAT	CGT	AAA	GAA	GAC	GGT	GTT	ATC	AAA	AAT	1776
	Gln	Met	Leu	Leu	Ile	Thr	Asp	Arg	Lys	Glu	Asp	Gly	Val	Ile	Lys	Asn	
				580					585					590			
80	AAC	TTA	GAG	AAT	TTG	TCT	GAT	CTG	TAT	TTG	GTT	AGT	TTG	CTG	GCC	CAG	1824
	Asn	Leu	Glu	Asn	Leu	Ser	Asp	Leu	Tyr	Leu	Val	Ser	Leu	Leu	Ala	Gln	

	595	600	605	
5	ATT CAT AAC CTG ACT ATT GCT GAA TTG AAC ATT TTG TTG GTG ATT TGT 1872 Ile His Asn Leu Thr Ile Ala Glu Leu Asn Ile Leu Leu Val Ile Cys 610 615 620			
10	GGC TAT GGC GAC ACC AAC ATT TAT CAG ATT ACC GAC GAT AAT TTA GCC 1920 Gly Tyr Gly Asp Thr Asn Ile Tyr Gln Ile Thr Asp Asp Asn Leu Ala 625 630 635 640			
15	AAA ATA GTG GAA ACA TTG TTG TGG ATC ACT CAA TGG TTG AAG ACC CAA 1968 Lys Ile Val Glu Thr Leu Leu Trp Ile Thr Gln Trp Leu Lys Thr Gln 645 650 655			
20	AAA TGG ACA GTT ACC GAC CTG TTT CTG ATG ACC ACG GCC ACT TAC AGC 2016 Lys Trp Thr Val Thr Asp Leu Phe Leu Met Thr Thr Ala Thr Tyr Ser 660 665 670			
25	ACC ACT TTA ACG CCA GAA ATT AGC AAT CTG ACG GCT ACG TTG TCT TCA 2064 Thr Thr Leu Thr Pro Glu Ile Ser Asn Leu Thr Ala Thr Leu Ser Ser 675 680 685			
30	ACT TTG CAT GGC AAA GAG AGT CTG ATT GGG GAA GAT CTG AAA AGA GCA 2112 Thr Leu His Gly Lys Glu Ser Leu Ile Gly Glu Asp Leu Lys Arg Ala 690 695 700			
35	ATG GCG CCT TGC TTC ACT TCG GCT TTG CAT TTG ACT TCT CAA GAA GTT 2160 Met Ala Pro Cys Phe Thr Ser Ala Leu His Leu Thr Ser Gln Glu Val 705 710 715 720			
40	GCG TAT GAC CTG CTG TTG TGG ATA GAC CAG ATT CAA CCG GCA CAA ATA 2208 Ala Tyr Asp Leu Leu Trp Ile Asp Gln Ile Gln Pro Ala Gln Ile 725 730 735			
45	ACT GTT GAT GGG TTT TGG GAA GAA GTG CAA ACA ACA CCA ACC AGC TTG 2256 Thr Val Asp Gly Phe Trp Glu Glu Val Gln Thr Thr Pro Thr Ser Leu 740 745 750			
50	AAG GTG ATT ACC TTT GCT CAG GTG CTG GCA CAA TTG AGC CTG ATC TAT 2304 Lys Val Ile Thr Phe Ala Gln Val Leu Ala Gln Leu Ser Leu Ile Tyr 755 760 765			
55	CGT CGT ATT GGG TTA AGT GAA ACG GAA CTG TCA CTG ATC GTG ACT CAA 2352 Arg Arg Ile Gly Leu Ser Glu Thr Glu Leu Ser Leu Ile Val Thr Gln 770 775 780			
60	TCT TCT CTG CTA GTG GCA GGC AAA AGC ATA CTG GAT CAC GGT CTG TTA 2400 Ser Ser Leu Leu Val Ala Gly Lys Ser Ile Leu Asp His Gly Leu Leu 785 790 795 800			
65	ACC CTG ATG GCC TTG GAA GGT TTT CAT ACC TGG GTT AAT GGC TTG GGG 2448 Thr Leu Met Ala Leu Glu Gly Phe His Thr Trp Val Asn Gly Leu Gly 805 810 815			
70	CAA CAT GCC TCC TTG ATA TTG GCG GCG TTG AAA GAC GGA GCC TTG ACA 2496 Gln His Ala Ser Leu Ile Leu Ala Ala Leu Lys Asp Gly Ala Leu Thr 820 825 830			
75	GTT ACC GAT GTA GCA CAA GCT ATG AAT AAG GAG GAA TCT CTC CTA CAA 2544 Val Thr Asp Val Ala Gln Ala Met Asn Lys Glu Glu Ser Leu Leu Gln 835 840 845			
80	ATG GCA GCT AAT CAG GTG GAG AAG GAT CTA ACA AAA CTG ACC AGT TGG 2592 Met Ala Ala Asn Gln Val Glu Lys Asp Leu Thr Lys Leu Thr Ser Trp 850 855 860			

	ACA	CAG	ATT	GAC	GCT	ATT	CTG	CAA	TGG	TTA	CAG	ATG	TCT	TCG	GCC	TTG	2640
	Thr	Gln	Ile	Asp	Ala	Ile	Leu	Gln	Trp	Leu	Gln	Met	Ser	Ser	Ala	Leu	
	865					870					875					880	
5	GCG	GTT	TCT	CCA	CTG	GAT	CTG	GCA	GGG	ATG	ATG	GCC	CTG	AAA	TAT	GGG	2688
	Ala	Val	Ser	Pro	Leu	Asp	Leu	Ala	Gly	Met	Met	Ala	Leu	Lys	Tyr	Gly	
					885					890					895		
10	ATA	GAT	CAT	AAC	TAT	GCT	GCC	TGG	CAA	GCT	GCG	GCG	GCT	GCG	CTG	ATG	2736
	Ile	Asp	His	Asn	Tyr	Ala	Ala	Trp	Gln	Ala	Ala	Ala	Ala	Ala	Leu	Met	
				900					905					910			
15	GCT	GAT	CAT	GCT	AAT	CAG	GCA	CAG	AAA	AAA	CTG	GAT	GAG	ACG	TTC	AGT	2784
	Ala	Asp	His	Ala	Asn	Gln	Ala	Gln	Lys	Lys	Leu	Asp	Glu	Thr	Phe	Ser	
			915					920					925				
20	AAG	GCA	TTA	TGT	AAC	TAT	TAT	ATT	AAT	GCT	GTT	GTC	GAT	AGT	GCT	GCT	2832
	Lys	Ala	Leu	Cys	Asn	Tyr	Tyr	Ile	Asn	Ala	Val	Val	Asp	Ser	Ala	Ala	
	930						935					940					
25	GGA	GTA	CGT	GAT	CGT	AAC	GGT	TTA	TAT	ACC	TAT	TTG	CTG	ATT	GAT	AAT	2880
	Gly	Val	Arg	Asp	Arg	Asn	Gly	Leu	Tyr	Thr	Tyr	Leu	Leu	Ile	Asp	Asn	
	945					950					955					960	
30	CAG	GTT	TCT	GCC	GAT	GTG	ATC	ACT	TCA	CGT	ATT	GCA	GAA	GCT	ATC	GCC	2928
	Gln	Val	Ser	Ala	Asp	Val	Ile	Thr	Ser	Arg	Ile	Ala	Glu	Ala	Ile	Ala	
					965					970					975		
35	GGT	ATT	CAA	CTG	TAC	GTT	AAC	CGG	GCT	TTA	AAC	CGA	GAT	GAA	GGT	CAG	2976
	Gly	Ile	Gln	Leu	Tyr	Val	Asn	Arg	Ala	Leu	Asn	Arg	Asp	Glu	Gly	Gln	
				980					985					990			
40	CTT	GCA	TCG	GAC	GTT	AGT	ACC	CGT	CAG	TTC	TTC	ACT	GAC	TGG	GAA	CGT	3024
	Leu	Ala	Ser	Asp	Val	Ser	Thr	Arg	Gln	Phe	Phe	Thr	Asp	Trp	Glu	Arg	
			995					1000					1005				
45	TAC	AAT	AAA	CGT	TAC	AGT	ACT	TGG	GCT	GGT	GTC	TCT	GAA	CTG	GTC	TAT	3072
	Tyr	Asn	Lys	Arg	Tyr	Ser	Thr	Trp	Ala	Gly	Val	Ser	Glu	Leu	Val	Tyr	
	1010						1015					1020					
50	TAT	CCA	GAA	AAC	TAT	GTT	GAT	CCC	ACT	CAG	CGC	ATT	GGG	CAA	ACC	AAA	3120
	Tyr	Pro	Glu	Asn	Tyr	Val	Asp	Pro	Thr	Gln	Arg	Ile	Gly	Gln	Thr	Lys	
	1025					1030					1035					1040	
55	ATG	ATG	GAT	GCG	CTG	TTG	CAA	TCC	ATC	AAC	CAG	AGC	CAG	CTA	AAT	GCG	3168
	Met	Met	Asp	Ala	Leu	Leu	Gln	Ser	Ile	Asn	Gln	Ser	Gln	Leu	Asn	Ala	
					1045					1050					1055		
60	GAT	ACG	GTG	GAA	GAT	GCT	TTC	AAA	ACT	TAT	TTG	ACC	AGC	TTT	GAG	CAG	3216
	Asp	Thr	Val	Glu	Asp	Ala	Phe	Lys	Thr	Tyr	Leu	Thr	Ser	Phe	Glu	Gln	
				1060					1065					1070			
65	GTA	GCA	AAT	CTG	AAA	GTA	ATT	AGT	GCT	TAC	CAC	GAT	AAT	GTG	AAT	GTG	3264
	Val	Ala	Asn	Leu	Lys	Val	Ile	Ser	Ala	Tyr	His	Asp	Asn	Val	Asn	Val	
			1075					1080					1085				
70	GAT	CAA	GGA	TTA	ACT	TAT	TTT	ATC	GGT	ATC	GAC	CAA	GCA	GCT	CCG	GGT	3312
	Asp	Gln	Gly	Leu	Thr	Tyr	Phe	Ile	Gly	Ile	Asp	Gln	Ala	Ala	Pro	Gly	
		1090					1095					1100					
75	ACG	TAT	TAC	TGG	CGT	AGT	GTT	GAT	CAC	AGC	AAA	TGT	GAA	AAT	GGC	AAG	3360
	Thr	Tyr	Tyr	Trp	Arg	Ser	Val	Asp	His	Ser	Lys	Cys	Glu	Asn	Gly	Lys	
	1105					1110					1115					1120	
80	TTT	GCC	GCT	AAT	GCT	TGG	GGT	GAG	TGG	AAT	AAA	ATT	ACC	TGT	GCT	GTC	3408
	Phe	Ala	Ala	Asn	Ala	Trp	Gly	Glu	Trp	Asn	Lys	Ile	Thr	Cys	Ala	Val	

	1125	1130	1135	
5	AAT CCT TGG AAA AAT ATC ATC CGT CCG GTT GTT TAT ATG TCC CGC TTA 3456 Asn Pro Trp Lys Asn Ile Ile Arg Pro Val Val Tyr Met Ser Arg Leu 1140 1145 1150			
10	TAT CTG CTA TGG CTG GAG CAG CAA TCA AAG AAA AGT GAT GAT GGT AAA 3504 Tyr Leu Leu Trp Leu Glu Gln Gln Ser Lys Lys Ser Asp Asp Gly Lys 1155 1160 1165			
15	ACC ACG ATT TAT CAA TAT AAC TTA AAA CTG GCT CAT ATT CGT TAC GAC 3552 Thr Thr Ile Tyr Gln Tyr Asn Leu Lys Leu Ala His Ile Arg Tyr Asp 1170 1175 1180			
20	GGT AGT TGG AAT ACA CCA TTT ACT TTT GAT GTG ACA GAA AAG GTA AAA 3600 Gly Ser Trp Asn Thr Pro Phe Thr Phe Asp Val Thr Glu Lys Val Lys 1185 1190 1195 1200			
25	AAT TAC ACG TCG AGT ACT GAT GCT GCT GAA TCT TTA GGG TTG TAT TGT 3648 Asn Tyr Thr Ser Ser Thr Asp Ala Ala Glu Ser Leu Gly Leu Tyr Cys 1205 1210 1215			
30	ACT GGT TAT CAA GGG GAA GAC ACT CTA TTA GTT ATG TTC TAT TCG ATG 3696 Thr Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Ser Met 1220 1225 1230			
35	CAG AGT AGT TAT AGC TCC TAT ACC GAT AAT AAT GCG CCG GTC ACT GGG 3744 Gln Ser Ser Tyr Ser Ser Tyr Thr Asp Asn Asn Ala Pro Val Thr Gly 1235 1240 1245			
40	CTA TAT ATT TTC GCT GAT ATG TCA TCA GAC AAT ATG ACG AAT GCA CAA 3792 Leu Tyr Ile Phe Ala Asp Met Ser Ser Asp Asn Met Thr Asn Ala Gln 1250 1255 1260			
45	GCA ACT AAC TAT TGG AAT AAC AGT TAT CCG CAA TTT GAT ACT GTG ATG 3840 Ala Thr Asn Tyr Trp Asn Asn Ser Tyr Pro Gln Phe Asp Thr Val Met 1265 1270 1275 1280			
50	GCA GAT CCG GAT AGC GAC AAT AAA AAA GTC ATA ACC AGA AGA GTT AAT 3888 Ala Asp Pro Asp Ser Asp Asn Lys Lys Val Ile Thr Arg Arg Val Asn 1285 1290 1295			
55	AAC CGT TAT GCG GAG GAT TAT GAA ATT CCT TCC TCT GTG ACA AGT AAC 3936 Asn Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Thr Ser Asn 1300 1305 1310			
60	AGT AAT TAT TCT TGG GGT GAT CAC AGT TTA ACC ATG CTT TAT GGT GGT 3984 Ser Asn Tyr Ser Trp Gly Asp His Ser Leu Thr Met Leu Tyr Gly Gly 1315 1320 1325			
65	AGT GTT CCT AAT ATT ACT TTT GAA TCG GCG GCA GAA GAT TTA AGG CTA 4032 Ser Val Pro Asn Ile Thr Phe Glu Ser Ala Ala Glu Asp Leu Arg Leu 1330 1335 1340			
70	TCT ACC AAT ATG GCA TTG AGT ATT ATT CAT AAT GGA TAT GCG GGA ACC 4080 Ser Thr Asn Met Ala Leu Ser Ile Ile His Asn Gly Tyr Ala Gly Thr 1345 1350 1355 1360			
75	CGC CGT ATA CAA TGT AAT CTT ATG AAA CAA TAC GCT TCA TTA GGT GAT 4128 Arg Arg Ile Gln Cys Asn Leu Met Lys Gln Tyr Ala Ser Leu Gly Asp 1365 1370 1375			
80	AAA TTT ATA ATT TAT GAT TCA TCA TTT GAT GAT GCA AAC CGT TTT AAT 4176 Lys Phe Ile Ile Tyr Asp Ser Ser Phe Asp Asp Ala Asn Arg Phe Asn 1380 1385 1390			

1
CTG GTG CCA TTG TTT AAA TTC GGA AAA GAC GAG AAC TCA GAT GAT AGT 4011
Leu Val Pro Leu Phe Lys Phe Gly Lys Asp Glu Asn Ser Asp Asp Ser
1395 1400 1405

5
ATT TGT ATA TAT AAT GAA AAC CCT TCC TCT GAA GAT AAG AAG TGG TAT 4271
Ile Cys Ile Tyr Asn Glu Asn Pro Ser Ser Glu Asp Lys Lys Trp Tyr
1410 1415 1420

10
TTT TCT TCG AAA GAT GAC AAT AAA ACA GCG GAT TAT AAT GGT GGA ACT 4321
Phe Ser Ser Lys Asp Asp Asn Lys Thr Ala Asp Tyr Asn Gly Gly Thr
1425 1430 1435 1440

15
CAA TGT ATA GAT GCT GGA ACC AGT AAC AAA GAT TTT TAT TAT AAT CTC 4361
Gln Cys Ile Asp Ala Gly Thr Ser Asn Lys Asp Phe Tyr Tyr Asn Leu
1445 1450 1455

20
CAG GAG ATT GAA GTA ATT AGT GTT ACT GGT GGG TAT TGG TCG AGT TAT 4411
Gln Glu Ile Glu Val Ile Ser Val Thr Gly Gly Tyr Trp Ser Ser Tyr
1460 1465 1470

25
AAA ATA TCC AAC CCG ATT AAT ATC AAT ACG GGC ATT GAT AGT GCT AAA 4461
Lys Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
1475 1480 1485

30
GTA AAA GTC ACC GTA AAA GCG GGT GGT GAC GAT CAA ATC TTT ACT GCT 4511
Val Lys Val Thr Val Lys Ala Gly Gly Asp Asp Gln Ile Phe Thr Ala
1490 1495 1500

35
GAT AAT AGT ACC TAT GTT CCT CAG CAA CCG GCA CCC AGT TTT GAG GAG 4561
Asp Asn Ser Thr Tyr Val Pro Gln Gln Pro Ala Pro Ser Phe Glu Glu
1505 1510 1515 1520

40
ATG ATT TAT CAG TTC AAT AAC CTG ACA ATA GAT TGT AAG AAT TTA AAT 4601
Met Ile Tyr Gln Phe Asn Asn Leu Thr Ile Asp Cys Lys Asn Leu Asn
1525 1530 1535

45
TTC ATC GAC AAT CAG GCA CAT ATT GAG ATT GAT TTC ACC GCT ACG GCA 4651
Phe Ile Asp Asn Gln Ala His Ile Glu Ile Asp Phe Thr Ala Thr Ala
1540 1545 1550

50
CAA GAT GGC CGA TTC TTG GGT GCA GAA ACT TTT ATT ATC CCG GTA ACT 4701
Gln Asp Gly Arg Phe Leu Gly Ala Glu Thr Phe Ile Ile Pro Val Thr
1555 1560 1565

55
AAA AAA GTT CTC GGT ACT GAG AAC GTG ATT GCG TTA TAT AGC GAA AAT 4752
Lys Lys Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn
1570 1575 1580

60
AAC GGT GTT CAA TAT ATG CAA ATT GGC GCA TAT CGT ACC CGT TTG AAT 4800
Asn Gly Val Gln Tyr Met Gln Ile Gly Ala Tyr Arg Thr Arg Leu Asn
1585 1590 1595 1600

65
ACG TTA TTC GCT CAA CAG TTG GTT AGC CGT GCT AAT CGT GGC ATT GAT 4848
Thr Leu Phe Ala Gln Gln Leu Val Ser Arg Ala Asn Arg Gly Ile Asp
1605 1610 1615

70
GCA GTG CTC AGT ATG GAA ACT CAG AAT ATT CAG GAA CCG CAA TTA GGA 4896
Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly
1620 1625 1630

75
GCG GGC ACA TAT GTG CAG CTT GTG TTG GAT AAA TAT GAT GAG TCT ATT 4944
Ala Gly Thr Tyr Val Gln Leu Val Leu Asp Lys Tyr Asp Glu Ser Ile
1635 1640 1645

80
CAT GGC ACT AAT AAA AGC TTT GCT ATT GAA TAT GTT GAT ATA TTT AAA 4992
His Gly Thr Asn Lys Ser Phe Ala Ile Glu Tyr Val Asp Ile Phe Lys

	1650	1655	1660	
5	GAG AAC GAT AGT TTT GTG ATT TAT CAA GGA GAA CTT AGC GAA ACA AGT 5040 Glu Asn Asp Ser Phe Val Ile Tyr Gln Gly Glu Leu Ser Glu Thr Ser 1665 1670 1675 1680			
10	CAA ACT GTT GTG AAA GTT TTC TTA TCC TAT TTT ATA GAG GCG ACT GGA 5088 Gln Thr Val Val Lys Val Phe Leu Ser Tyr Phe Ile Glu Ala Thr Gly 1685 1690 1695			
15	AAT AAG AAC CAC TTA TGG GTA CGT GCT AAA TAC CAA AAG GAA ACG ACT 5136 Asn Lys Asn His Leu Trp Val Arg Ala Lys Tyr Gln Lys Glu Thr Thr 1700 1705 1710			
20	GAT AAG ATC TTG TTC GAC CGT ACT GAT GAG AAA GAT CCG CAC GGT TGG 5184 Asp Lys Ile Leu Phe Asp Arg Thr Asp Glu Lys Asp Pro His Gly Trp 1715 1720 1725			
25	TTT CTC AGC GAC GAT CAC AAG ACC TTT AGT GGT CTC TCT TCC GCA CAG 5232 Phe Leu Ser Asp Asp His Lys Thr Phe Ser Gly Leu Ser Ser Ala Gln 1730 1735 1740			
30	GCA TTA AAG AAC GAC AGT GAA CCG ATG GAT TTC TCT GGC GCC AAT GCT 5280 Ala Leu Lys Asn Asp Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ala 1745 1750 1755 1760			
35	CTC TAT TTC TGG GAA CTG TTC TAT TAC ACG CCG ATG ATG ATG GCT CAT 5328 Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Met Met Ala His 1765 1770 1775			
40	CGT TTG TTG CAG GAA CAG AAT TTT GAT GCG GCG AAC CAT TGG TTC CGT 5376 Arg Leu Leu Gln Glu Asn Phe Asp Ala Ala Asn His Trp Phe Arg 1780 1785 1790			
45	TAT GTC TGG AGT CCA TCC GGT TAT ATC GTT GAT GGT AAA ATT GCT ATC 5424 Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Asp Gly Lys Ile Ala Ile 1795 1800 1805			
50	TAC CAC TGG AAC GTG CGA CCG CTG GAA GAA GAC ACC AGT TGG AAT GCA 5472 Tyr His Trp Asn Val Arg Pro Leu Glu Glu Asp Thr Ser Trp Asn Ala 1810 1815 1820			
55	CAA CAA CTG GAC TCC ACC GAT CCA GAT GCT GTA GCC CAA GAT GAT CCG 5520 Gln Gln Leu Asp Ser Thr Asp Pro Asp Ala Val Ala Gln Asp Asp Pro 1825 1830 1835 1840			
60	ATG CAC TAC AAG GTG GCT ACC TTT ATG GCG ACG TTG GAT CTG CTA ATG 5563 Met His Tyr Lys Val Ala Thr Phe Met Ala Thr Leu Asp Leu Leu Met 1845 1850 1855			
65	GCC CGT GGT GAT GCT GCT TAC CGC CAG TTA GAG CGT GAT ACG TTG GCT 5616 Ala Arg Gly Asp Ala Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Ala 1860 1865 1870			
70	GAA GCT AAA ATG TGG TAT ACA CAG GCG CTT AAT CTG TTG GGT GAT GAG 5664 Glu Ala Lys Met Trp Tyr Thr Gln Ala Leu Asn Leu Leu Gly Asp Glu 1875 1880 1885			
75	CCA CAA GTG ATG CTG AGT ACG ACT TGG GCT AAT CCA ACA TTG GGT AAT 5712 Pro Gln Val Met Leu Ser Thr Thr Trp Ala Asn Pro Thr Leu Gly Asn 1890 1895 1900			
80	GCT GCT TCA AAA ACC ACA CAG CAG GTT CGT CAG CAA GTG CTT ACC CAG 5760 Ala Ala Ser Lys Thr Thr Gln Gln Val Arg Gln Gln Val Leu Thr Gln 1905 1910 1915 1920			

	TTG CGT CTC AAT AGC AGG GTA AAA ACC CCG TTG CTA GGA ACA GCC AAT 5408 Leu Arg Leu Asn Ser Arg Val Lys Thr Pro Leu Leu Gly Thr Ala Asn 1925 1930 1935
5	TCC CTG ACC GCT TTA TTC CTG CCG CAG GAA AAT AGC AAG CTC AAA GGC 5356 Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn Ser Lys Leu Lys Gly 1940 1945 1950
10	TAC TGG CGG ACA CTG GCG CAG CGT ATG TTT AAT TTA CGT CAT AAT CTG 5904 Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn Leu Arg His Asn Leu 1955 1960 1965
15	TCG ATT GAC GGC CAG CCG CTC TCC TTG CCG CTG TAT GCT AAA CCG GCT 5352 Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu Tyr Ala Lys Pro Ala 1970 1975 1980
20	GAT CCA AAA GCT TTA CTG AGT GCG GCG GTT TCA GCT TCT CAA GGG GGA 6000 Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser Ala Ser Gln Gly Gly 1985 1990 1995 2000
	GCC GAC TTG CCG AAG GCG CCG CTG ACT ATT CAC CGC TTC CCT CAA ATG 6048 Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His Arg Phe Pro Gln Met 2005 2010 2015
25	CTA GAA GGG GCA CGG GGC TTG GTT AAC CAG CTT ATA CAG TTC GGT AGT 6096 Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu Ile Gln Phe Gly Ser 2020 2025 2030
30	TCA CTA TTG GGG TAC AGT GAG CGT CAG GAT GCG GAA GCT ATG AGT CAA 6144 Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala Glu Ala Met Ser Gln 2035 2040 2045
35	CTA CTG CAA ACC CAA GCC AGC GAG TTA ATA CTG ACC AGT ATT CGT ATG 6192 Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu Thr Ser Ile Arg Met 2050 2055 2060
40	CAG GAT AAC CAA TTG GCA GAG CTG GAT TCG GAA AAA ACC GCC TTG CAA 6240 Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu Lys Thr Ala Leu Gln 2055 2070 2075 2080
	GTC TCT TTA GCT GGA GTG CAA CAA CGG TTT GAC AGC TAT AGC CAA CTG 6288 Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp Ser Tyr Ser Gln Leu 2085 2090 2095
45	TAT GAG GAG AAC ATC AAC GCA GGT GAG CAG CGA GCG CTG GCG TTA CGC 6336 Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg Ala Leu Ala Leu Arg 2100 2105 2110
50	TCA GAA TCT GCT ATT GAG TCT CAG GGA GCG CAG ATT TCC CGT ATG GCA 6384 Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln Ile Ser Arg Met Ala 2115 2120 2125
55	GGC GCG GGT GTT GAT ATG GCA CCA AAT ATC TTC GGC CTG GCT GAT GGC 6432 Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe Gly Leu Ala Asp Gly 2130 2135 2140
60	GGC ATG CAT TAT GGT GCT ATT GCC TAT GCC ATC GCT GAC GGT ATT GAG 6480 Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile Ala Asp Gly Ile Glu 2145 2150 2155 2160
	TTG AGT GCT TCT GCC AAG ATG GTT GAT GCG GAG AAA GTT GCT CAG TCG 6528 Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu Lys Val Ala Gln S r 2165 2170 2175
65	GAA ATA TAT CGC CGT CGC CGT CAA GAA TGG AAA ATT CAG CGT GAC AAC 6576 Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys Ile Gln Arg Asp Asn

	2180	2185	2190	
5	GCA CAA GCG GAG ATT AAC CAG TTA AAC GCG CAA CTG GAA TCA CTG TCT 6624 Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln Leu Glu Ser Leu Ser 2195 2200 2205			
10	ATT CGC CGT GAA GCC GCT GAA ATG CAA AAA GAG TAC CTG AAA ACC CAG 6672 Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu Tyr Leu Lys Thr Gln 2210 2215 2220			
15	CAA GCT CAG GCG CAG GCA CAA CTT ACT TTC TTA AGA AGC AAA TTC AGT 6720 Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu Arg Ser Lys Phe Ser 2225 2230 2235 2240			
20	AAT CAA GCG TTA TAT AGT TGG TTA CGA GGG CGT TTG TCA GGT ATT TAT 6768 Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg Leu Ser Gly Ile Tyr 2245 2250 2255			
25	TTC CAG TTC TAT GAC TTG GCC GTA TCA CGT TGC CTG ATG GCA GAG CAA 6816 Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys Leu Met Ala Glu Gln 2260 2265 2270			
30	TCC TAT CAA TGG GAA GCT AAT GAT AAT TCC ATT AGC TTT GTC AAA CCG 6864 Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile Ser Phe Val Lys Pro 2275 2280 2285			
35	GGT GCA TGG CAA GGA ACT TAC GCC GGC TTA TTG TGT GGA GAA GCT TTG 6912 Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu Cys Gly Glu Ala Leu 2290 2295 2300			
40	ATA CAA AAT CTG GCA CAA ATG GAA GAG GCA TAT CTG AAA TGG GAA TCT 6960 Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr Leu Lys Trp Glu Ser 2305 2310 2315 2320			
45	CGC GCT TTG GAA GTA GAA CGC ACG GTT TCA TTG GCA GTG GTT TAT GAT 7008 Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Val Val Tyr Asp 2325 2330 2335			
50	TCA CTG GAA GGT AAT GAT CGT TTT AAT TTA GCG GAA CAA ATA CCT GCA 7056 Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala Glu Gln Ile Pro Ala 2340 2345 2350			
55	TTA TTG GAT AAG GGG GAG GGA ACA GCA GGA ACT AAA GAA AAT GGG TTA 7104 Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr Lys Glu Asn Gly Leu 2355 2360 2365			
60	TCA TTG GCT AAT GCT ATC CTG TCA GCT TCG GTC AAA TTG TCC GAC TTG 7152 Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val Lys Leu Ser Asp Leu 2370 2375 2380			
65	AAA CTG GGA ACG GAT TAT CCA GAC AGT ATC GTT GGT AGC AAC AAG GTT 7200 Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val Gly Ser Asn Lys Val 2385 2390 2395 2400			
70	CGT CGT ATT AAG CAA ATC AGT GTT TCG CTA CCT GCA TTG GTT GGG CCT 7248 Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro Ala Leu Val Gly Pro 2405 2410 2415			
75	TAT CAG GAT GTT CAG GCT ATG CTC AGC TAT GGT GGC AGT ACT CAA TTG 7296 Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly Gly Ser Thr Gln Leu 2420 2425 2430			
80	CCG AAA GGT TGT TCA GCG TTG GCT GTG TCT CAT GGT ACC AAT GAT AGT 7344 Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser 2435 2440 2445			

GGT CAG TTC CAG TTG GAT TTC AAT GAC GGC AAA TAC CTG CCA TTT GAA 7392
 Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Tyr Leu Pro Phe Glu
 2450 2455 2460
 5 GGT ATT GCT CTT GAT GAT CAG GGT ACA CTG AAT CTT CAA TTT CCG AAT 7440
 Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn Leu Gln Phe Pro Asn
 2465 2470 2475 2480
 10 GCT ACC GAC AAG CAG AAA GCA ATA TTG CAA ACT ATG AGC GAT ATT ATT 7488
 Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr Met Ser Asp Ile Ile
 2485 2490 2495
 TTG CAT ATT CGT TAT ACC ATC CGT TAA 7515
 Leu His Ile Arg Tyr Thr Ile Arg *
 15 2500 2505

(2) INFORMATION FOR SEQ ID NO:12:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2505 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

30 Met Gln Asn Ser Leu Ser Ser Thr Ile Asp Thr Ile Cys Gln Lys Leu
 1 5 10 15
 Gln Leu Thr Cys Pro Ala Glu Ile Ala Leu Tyr Pro Phe Asp Thr Phe
 20 25 30
 35 Arg Glu Lys Thr Arg Gly Met Val Asn Trp Gly Glu Ala Lys Arg Ile
 35 40 45
 Tyr Glu Ile Ala Gln Ala Glu Gln Asp Arg Asn Leu Leu His Glu Lys
 40 50 55 60
 Arg Ile Phe Ala Tyr Ala Asn Pro Leu Leu Lys Asn Ala Val Arg Leu
 65 70 75 80
 45 Gly Thr Arg Gln Met Leu Gly Phe Ile Gln Gly Tyr Ser Asp Leu Phe
 85 90 95
 Gly Asn Arg Ala Asp Asn Tyr Ala Ala Pro Gly Ser Val Ala Ser Met
 100 105 110
 50 Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Lys Asn
 115 120 125
 Leu His Asp Ser Ser Ser Ile Tyr Tyr Leu Asp Lys Arg Arg Pro Asp
 55 130 135 140
 Leu Ala Ser Leu Met Leu Ser Gln Lys Asn Met Asp Glu Glu Ile Ser
 145 150 155 160
 60 Thr Leu Ala Leu Ser Asn Glu Leu Cys Leu Ala Gly Ile Glu Thr Lys
 165 170 175
 Thr Gly Lys Ser Gln Asp Glu Val Met Asp Met Leu Ser Thr Tyr Arg
 180 185 190
 65

	Leu Ser Gly Glu Thr Pro Tyr His His Ala Tyr Glu Thr Val Arg Glu	195	200	205
5	Ile Val His Glu Arg Asp Pro Gly Phe Arg His Leu Ser Gln Ala Pro	210	215	220
	Ile Val Ala Ala Lys Leu Asp Pro Val Thr Leu Leu Gly Ile Ser Ser	225	230	235
10	His Ile Ser Pro Glu Leu Tyr Asn Leu Leu Ile Glu Glu Ile Pro Glu	245	250	255
	Lys Asp Glu Ala Ala Leu Asp Thr Leu Tyr Lys Thr Asn Phe Gly Asp	260	265	270
15	Ile Thr Thr Ala Gln Leu Met Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr	275	280	285
20	Gly Val Ser Pro Glu Asp Ile Ala Tyr Val Thr Thr Ser Leu Ser His	290	295	300
	Val Gly Tyr Ser Ser Asp Ile Leu Val Ile Pro Leu Val Asp Gly Val	305	310	315
25	Gly Lys Met Glu Val Val Arg Val Thr Arg Thr Pro Ser Asp Asn Tyr	325	330	335
	Thr Ser Gln Thr Asn Tyr Ile Glu Leu Tyr Pro Gln Gly Gly Asp Asn	340	345	350
30	Tyr Leu Ile Lys Tyr Asn Leu Ser Asn Ser Phe Gly Leu Asp Asp Phe	355	360	365
35	Tyr Leu Gln Tyr Lys Asp Gly Ser Ala Asp Trp Thr Glu Ile Ala His	370	375	380
	Asn Pro Tyr Pro Asp Met Val Ile Asn Gln Lys Tyr Glu Ser Gln Ala	385	390	395
40	Thr Ile Lys Arg Ser Asp Ser Asp Asn Ile Leu Ser Ile Gly Leu Gln	405	410	415
	Arg Trp His Ser Gly Ser Tyr Asn Phe Ala Ala Ala Asn Phe Lys Ile	420	425	430
45	Asp Gln Tyr Ser Pro Lys Ala Phe Leu Leu Lys Met Asn Lys Ala Ile	435	440	445
50	Arg Leu Leu Lys Ala Thr Gly Leu Ser Phe Ala Thr Leu Glu Arg Ile	450	455	460
	Val Asp Ser Val Asn Ser Thr Lys Ser Ile Thr Val Glu Val Leu Asn	465	470	475
55	Lys Val Tyr Arg Val Lys Phe Tyr Ile Asp Arg Tyr Gly Ile Ser Glu	485	490	495
	Glu Thr Ala Ala Ile Leu Ala Asn Ile Asn Ile Ser Gln Gln Ala Val	500	505	510
60	Gly Asn Gln Leu Ser Gln Phe Glu Gln Leu Phe Asn His Pro Pro Leu	515	520	525
65	Asn Gly Ile Arg Tyr Glu Ile Ser Glu Asp Asn Ser Lys His Leu Pro	530	535	540

	Asn	Pro	Asp	Leu	Asn	Leu	Lys	Pro	Asp	Ser	Thr	Gly	Asp	Asp	Gln	Arg	
	545					550					555					560	
5	Lys	Ala	Val	Leu	Lys	Arg	Ala	Phe	Gln	Val	Asn	Ala	Ser	Glu	Leu	Tyr	
					565					570					575		
	Gln	Met	Leu	Leu	Ile	Thr	Asp	Arg	Lys	Glu	Asp	Gly	Val	Ile	Lys	Asn	
				580					585					590			
10	Asn	Leu	Glu	Asn	Leu	Ser	Asp	Leu	Tyr	Leu	Val	Ser	Leu	Leu	Ala	Gln	
			595					600					605				
	Ile	His	Asn	Leu	Thr	Ile	Ala	Glu	Leu	Asn	Ile	Leu	Leu	Val	Ile	Cys	
15		610					615					620					
	Gly	Tyr	Gly	Asp	Thr	Asn	Ile	Tyr	Gln	Ile	Thr	Asp	Asp	Asn	Leu	Ala	
	625					630					635					640	
20	Lys	Ile	Val	Glu	Thr	Leu	Leu	Trp	Ile	Thr	Gln	Trp	Leu	Lys	Thr	Gln	
					645					650					655		
	Lys	Trp	Thr	Val	Thr	Asp	Leu	Phe	Leu	Met	Thr	Thr	Ala	Thr	Tyr	Ser	
				660					665					670			
25	Thr	Thr	Leu	Thr	Pro	Glu	Ile	Ser	Asn	Leu	Thr	Ala	Thr	Leu	Ser	Ser	
			675					680					685				
	Thr	Leu	His	Gly	Lys	Glu	Ser	Leu	Ile	Gly	Glu	Asp	Leu	Lys	Arg	Ala	
30		690					695					700					
	Met	Ala	Pro	Cys	Phe	Thr	Ser	Ala	Leu	His	Leu	Thr	Ser	Gln	Glu	Val	
	705					710					715					720	
35	Ala	Tyr	Asp	Leu	Leu	Leu	Trp	Ile	Asp	Gln	Ile	Gln	Pro	Ala	Gln	Ile	
				725					730						735		
	Thr	Val	Asp	Gly	Phe	Trp	Glu	Glu	Val	Gln	Thr	Thr	Pro	Thr	Ser	Leu	
				740				745						750			
40	Lys	Val	Ile	Thr	Phe	Ala	Gln	Val	Leu	Ala	Gln	Leu	Ser	Leu	Ile	Tyr	
		755					760					765					
	Arg	Arg	Ile	Gly	Leu	Ser	Glu	Thr	Glu	Leu	Ser	Leu	Ile	Val	Thr	Gln	
45		770					775					780					
	Ser	Ser	Leu	Leu	Val	Ala	Gly	Lys	Ser	Ile	Leu	Asp	His	Gly	Leu	Leu	
	785				790						795					800	
50	Thr	Leu	Met	Ala	Leu	Glu	Gly	Phe	His	Thr	Trp	Val	Asn	Gly	Leu	Gly	
					805					810					815		
	Gln	His	Ala	Ser	Leu	Ile	Leu	Ala	Ala	Leu	Lys	Asp	Gly	Ala	Leu	Thr	
				820				825						830			
55	Val	Thr	Asp	Val	Ala	Gln	Ala	Met	Asn	Lys	Glu	Glu	Ser	Leu	Leu	Gln	
			835				840						845				
	Met	Ala	Ala	Asn	Gln	Val	Glu	Lys	Asp	Leu	Thr	Lys	Leu	Thr	Ser	Trp	
60		850					855					860					
	Thr	Gln	Ile	Asp	Ala	Ile	Leu	Gln	Trp	Leu	Gln	Met	Ser	Ser	Ala	Leu	
	865					870					875					880	
65	Ala	Val	Ser	Pro	Leu	Asp	Leu	Ala	Gly	Met	Met	Ala	Leu	Lys	Tyr	Gly	
					885				890						895		

	Ile Asp His Asn Tyr Ala Ala Trp Gln Ala Ala Ala Ala Ala Leu Met	900	905	910
5	Ala Asp His Ala Asn Gln Ala Gln Lys Lys Leu Asp Glu Thr Phe Ser	915	920	925
	Lys Ala Leu Cys Asn Tyr Tyr Ile Asn Ala Val Val Asp Ser Ala Ala	930	935	940
10	Gly Val Arg Asp Arg Asn Gly Leu Tyr Thr Tyr Leu Leu Ile Asp Asn	945	950	955
	Gln Val Ser Ala Asp Val Ile Thr Ser Arg Ile Ala Glu Ala Ile Ala	965	970	975
15	Gly Ile Gln Leu Tyr Val Asn Arg Ala Leu Asn Arg Asp Glu Gly Gln	980	985	990
20	Leu Ala Ser Asp Val Ser Thr Arg Gln Phe Phe Thr Asp Trp Glu Arg	995	1000	1005
	Tyr Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Glu Leu Val Tyr	1010	1015	1020
25	Tyr Pro Glu Asn Tyr Val Asp Pro Thr Gln Arg Ile Gly Gln Thr Lys	1025	1030	1035
	Met Met Asp Ala Leu Leu Gln Ser Ile Asn Gln Ser Gln Leu Asn Ala	1045	1050	1055
30	Asp Thr Val Glu Asp Ala Phe Lys Thr Tyr Leu Thr Ser Phe Glu Gln	1060	1065	1070
35	Val Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Val Asn Val	1075	1080	1085
	Asp Gln Gly Leu Thr Tyr Phe Ile Gly Ile Asp Gln Ala Ala Pro Gly	1090	1095	1100
40	Thr Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Cys Glu Asn Gly Lys	1105	1110	1115
	Phe Ala Ala Asn Ala Trp Gly Glu Trp Asn Lys Ile Thr Cys Ala Val	1125	1130	1135
45	Asn Pro Trp Lys Asn Ile Ile Arg Pro Val Val Tyr Met Ser Arg Leu	1140	1145	1150
50	Tyr Leu Leu Trp Leu Glu Gln Gln Ser Lys Lys Ser Asp Asp Gly Lys	1155	1160	1165
	Thr Thr Ile Tyr Gln Tyr Asn Leu Lys Leu Ala His Ile Arg Tyr Asp	1170	1175	1180
55	Gly Ser Trp Asn Thr Pro Phe Thr Phe Asp Val Thr Glu Lys Val Lys	1185	1190	1195
	Asn Tyr Thr Ser Ser Thr Asp Ala Ala Glu Ser Leu Gly Leu Tyr Cys	1205	1210	1215
60	Thr Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Ser Met	1220	1225	1230
65	Gln Ser Ser Tyr Ser Ser Tyr Thr Asp Asn Asn Ala Pro Val Thr Gly	1235	1240	1245

Leu Tyr Ile Phe Ala Asp Met Ser Ser Asp Asn Met Thr Asn Ala Gln
 1250 1255 1260
 5 Ala Thr Asn Tyr Trp Asn Asn Ser Tyr Pro Gln Phe Asp Thr Val Met
 1265 1270 1275 1280
 Ala Asp Pro Asp Ser Asp Asn Lys Lys Val Ile Thr Arg Arg Val Asn
 1285 1290 1295
 10 Asn Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Thr Ser Asn
 1300 1305 1310
 Ser Asn Tyr Ser Trp Gly Asp His Ser Leu Thr Met Leu Tyr Gly Gly
 1315 1320 1325
 15 Ser Val Pro Asn Ile Thr Phe Glu Ser Ala Ala Glu Asp Leu Arg Leu
 1330 1335 1340
 20 Ser Thr Asn Met Ala Leu Ser Ile Ile His Asn Gly Tyr Ala Gly Thr
 1345 1350 1355 1360
 Arg Arg Ile Gln Cys Asn Leu Met Lys Gln Tyr Ala Ser Leu Gly Asp
 1365 1370 1375
 25 Lys Phe Ile Ile Tyr Asp Ser Ser Phe Asp Asp Ala Asn Arg Phe Asn
 1380 1385 1390
 Leu Val Pro Leu Phe Lys Phe Gly Lys Asp Glu Asn Ser Asp Asp Ser
 1395 1400 1405
 30 Ile Cys Ile Tyr Asn Glu Asn Pro Ser Ser Glu Asp Lys Lys Trp Tyr
 1410 1415 1420
 35 Phe Ser Ser Lys Asp Asp Asn Lys Thr Ala Asp Tyr Asn Gly Gly Thr
 1425 1430 1435 1440
 Gln Cys Ile Asp Ala Gly Thr Ser Asn Lys Asp Phe Tyr Tyr Asn Leu
 1445 1450 1455
 40 Gln Glu Ile Glu Val Ile Ser Val Thr Gly Gly Tyr Trp Ser Ser Tyr
 1460 1465 1470
 Lys Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
 1475 1480 1485
 45 Val Lys Val Thr Val Lys Ala Gly Gly Asp Asp Gln Ile Phe Thr Ala
 1490 1495 1500
 50 Asp Asn Ser Thr Tyr Val Pro Gln Gln Pro Ala Pro Ser Phe Glu Glu
 1505 1510 1515 1520
 Met Ile Tyr Gln Phe Asn Asn Leu Thr Ile Asp Cys Lys Asn Leu Asn
 1525 1530 1535
 55 Phe Ile Asp Asn Gln Ala His Ile Glu Ile Asp Phe Thr Ala Thr Ala
 1540 1545 1550
 Gln Asp Gly Arg Phe Leu Gly Ala Glu Thr Phe Ile Ile Pro Val Thr
 1555 1560 1565
 60 Lys Lys Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn
 1570 1575 1580
 65 Asn Gly Val Gln Tyr Met Gln Ile Gly Ala Tyr Arg Thr Arg Leu Asn
 1585 1590 1595 1600

Thr Leu Phe Ala Gln Gln Leu Val Ser Arg Ala Asn Arg Gly Ile Asp
 1605 1610 1615
 5 Ala Val L u Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly
 1620 1625 1630
 Ala Gly Thr Tyr Val Gln Leu Val Leu Asp Lys Tyr Asp Glu Ser Ile
 1635 1640 1645
 10 His Gly Thr Asn Lys Ser Phe Ala Ile Glu Tyr Val Asp Ile Phe Lys
 1650 1655 1660
 Glu Asn Asp Ser Phe Val Ile Tyr Gln Gly Glu Leu Ser Glu Thr Ser
 1665 1670 1675 1680
 15 Gln Thr Val Val Lys Val Phe Leu Ser Tyr Phe Ile Glu Ala Thr Gly
 1685 1690 1695
 Asn Lys Asn His Leu Trp Val Arg Ala Lys Tyr Gln Lys Glu Thr Thr
 1700 1705 1710
 Asp Lys Ile Leu Phe Asp Arg Thr Asp Glu Lys Asp Pro His Gly Trp
 1715 1720 1725
 25 Phe Leu Ser Asp Asp His Lys Thr Phe Ser Gly Leu Ser Ser Ala Gln
 1730 1735 1740
 Ala Leu Lys Asn Asp Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ala
 1745 1750 1755 1760
 30 Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Met Met Ala His
 1765 1770 1775
 Arg Leu Leu Gln Glu Gln Asn Phe Asp Ala Ala Asn His Trp Phe Arg
 1780 1785 1790
 Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Asp Gly Lys Ile Ala Ile
 1795 1800 1805
 40 Tyr His Trp Asn Val Arg Pro Leu Glu Glu Asp Thr Ser Trp Asn Ala
 1810 1815 1820
 Gln Gln Leu Asp Ser Thr Asp Pro Asp Ala Val Ala Gln Asp Asp Pro
 1825 1830 1835 1840
 45 Met His Tyr Lys Val Ala Thr Phe Met Ala Thr Leu Asp Leu Leu Met
 1845 1850 1855
 Ala Arg Gly Asp Ala Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Ala
 1860 1865 1870
 50 Glu Ala Lys Met Trp Tyr Thr Gln Ala Leu Asn Leu Leu Gly Asp Glu
 1875 1880 1885
 Pro Gln Val Met Leu Ser Thr Thr Trp Ala Asn Pro Thr Leu Gly Asn
 1890 1895 1900
 Ala Ala Ser Lys Thr Thr Gln Gln Val Arg Gln Gln Val Leu Thr Gln
 1905 1910 1915 1920
 60 Leu Arg Leu Asn Ser Arg Val Lys Thr Pro Leu Leu Gly Thr Ala Asn
 1925 1930 1935
 Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn Ser Lys Leu Lys Gly
 1940 1945 1950

Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn Leu Arg His Asn Leu
 1955 1960 1965
 5 Ser Il Asp Gly Gln Pro Leu Ser Leu Pro Leu Tyr Ala Lys Pro Ala
 1970 1975 1980
 Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser Ala Ser Gln Gly Gly
 1985 1990 1995 2000
 10 Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His Arg Phe Pro Gln Met
 2005 2010 2015
 Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu Ile Gln Phe Gly Ser
 2020 2025 2030
 15 Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala Glu Ala Met Ser Gln
 2035 2040 2045
 Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu Thr Ser Ile Arg Met
 2050 2055 2060
 Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu Lys Thr Ala Leu Gln
 2065 2070 2075 2080
 25 Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp Ser Tyr Ser Gln Leu
 2085 2090 2095
 Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg Ala Leu Ala Leu Arg
 2100 2105 2110
 30 Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln Ile Ser Arg Met Ala
 2115 2120 2125
 Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe Gly Leu Ala Asp Gly
 2130 2135 2140
 Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile Ala Asp Gly Ile Glu
 2145 2150 2155 2160
 40 Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu Lys Val Ala Gln Ser
 2165 2170 2175
 Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys Ile Gln Arg Asp Asn
 2180 2185 2190
 45 Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln Leu Glu Ser Leu Ser
 2195 2200 2205
 Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu Tyr Leu Lys Thr Gln
 2210 2215 2220
 Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu Arg Ser Lys Phe Ser
 2225 2230 2235 2240
 55 Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg Leu Ser Gly Ile Tyr
 2245 2250 2255
 Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys Leu Met Ala Glu Gln
 2260 2265 2270
 60 Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile Ser Phe Val Lys Pro
 2275 2280 2285
 Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu Cys Gly Glu Ala L u
 2290 2295 2300
 65

Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr Leu Lys Trp Glu Ser
 2305 2310 2315 2320
 5 Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Val Val Tyr Asp
 2325 2330 2335
 Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala Glu Gln Ile Pro Ala
 2340 2345 2350
 10 Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr Lys Glu Asn Gly Leu
 2355 2360 2365
 Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val Lys Leu Ser Asp Leu
 2370 2375 2380
 15 Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val Gly Ser Asn Lys Val
 2385 2390 2395 2400
 20 Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro Ala Leu Val Gly Pro
 2405 2410 2415
 Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly Gly Ser Thr Gln Leu
 2420 2425 2430
 25 Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser
 2435 2440 2445
 Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Tyr Leu Pro Phe Glu
 2450 2455 2460
 30 Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn Leu Gln Phe Pro Asn
 2465 2470 2475 2480
 35 Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr Met Ser Asp Ile Ile
 2485 2490 2495
 Leu His Ile Arg Tyr Thr Ile Arg *
 2500 2505

40

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 45 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

55 Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Xaa Ala
 1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met Gln Asn Ser Gln Thr Phe Ser Val Gly Glu Leu
1 5 10

10 (2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
15 (A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Ala Gln Asp Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr
25 1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
35 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

40 Met Gln Asn Ser Leu
1 5

45 (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
50 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ala Phe Asn Ile Asp Asp Val Ser Leu Phe
60 1 5 10

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile Gly Ser
 1 5 10 15

Leu Gln Leu Phe Ile
 20

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Tyr Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro
 1 5 10

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Gly Ile Asp Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro
1 5 10 15

10 Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu
20 25

(2) INFORMATION FOR SEQ ID NO:22:

15

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
1 5 10 15

30

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 35 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

45 Thr Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:24:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 55 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

60

Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly

1	5	10	15
Val	Gln	Tyr	Met
			Gln
			Ile
			20

5

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6005 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: RBS
(B) LOCATION: 1..9

(ix) **FEATURE:**

- ```
(A) NAME/KEY: CDS
(B) LOCATION: 16..3585
(D) OTHER INFORMATION: /product= "P8"
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

|    |             |             |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-------------|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 30 | AAGAAGGAAT  | Met         | 1   | ATG | Ser | 5   | TCT | Glu | 10  | TCT | Leu | 15  | TTT | Phe | 20  | ACA | Thr | 25  | CAA | Gln | 30  | ACG | Thr | 35  | TTG | Leu | 40  | AAA | Lys | 45  | GAA | Glu | 50  |     |     |     |     |     |     |     |     |     |     |     |     |
| 35 | GCG CGC CGT | Ala Arg Arg | 15  | GAT | Asp | 20  | GCA | Ala | 25  | TTG | Leu | 30  | GTT | Val | 35  | GCT | Ala | 40  | CAT | His | 45  | TAT | Tyr | 50  | ATT | Ile | 55  | GCT | Ala | 60  | ACT | Thr | 65  | CAG | Gln | 70  | GTG | Val | 75  | CCC | Pro | 80  |     |     |     |
| 40 | GCA GAT TTA | Ala Asp Leu | 30  | AAA | Lys | 35  | GAG | Glu | 40  | AGT | Ser | 45  | ATC | Ile | 50  | CAG | Gln | 55  | ACC | Thr | 60  | GCG | Ala | 65  | GAT | Asp | 70  | CTG | Leu | 75  | CTC | Tyr | 80  | TAC | Glu | 85  | GAA | Tyr | 90  | TAT | Tyr | 95  |     |     |     |
| 45 | CTG TTG CTG | Leu Leu Leu | 45  | GAT | Asp | 50  | ACC | Thr | 55  | AAA | Lys | 60  | ATT | Ile | 65  | AGC | Ser | 70  | GAT | Asp | 75  | CTG | Leu | 80  | GTT | Val | 85  | ACT | Thr | 90  | ACT | Thr | 95  | TCA | Ser | 100 | CCG | Pro | 105 | CTG | Leu | 110 |     |     |     |
| 50 | TCC GAA GCG | Ser Glu Ala | 65  | ATT | Ile | 70  | GGC | Gly | 75  | AGT | Ser | 80  | CTG | Leu | 85  | CAA | Gln | 90  | TTG | Leu | 95  | TTT | Phe | 100 | ATT | Ile | 105 | CAT | His | 110 | CGT | Arg | 115 | GCG | Ala | 120 | ATA | Ile | 125 | GAG | Glu | 130 |     |     |     |
| 55 | GGC TAT GAC | Gly Tyr Asp | 80  | GGC | Gly | 85  | ACG | Thr | 90  | CTG | Leu | 95  | GCA | Ala | 100 | GAC | Asp | 105 | TCA | Ser | 110 | GCA | Ala | 115 | AAA | Lys | 120 | CCC | Pro | 125 | TAT | Tyr | 130 | TTT | Phe | 135 | GCC | Ala | 140 | GGG | Gly | 145 | GAT | Asp | 150 |
| 60 | GAA CAG TTT | Glu Gln Phe | 95  | TTA | Leu | 100 | TAT | Tyr | 105 | AAC | Asn | 110 | TGG | Trp | 115 | GAT | Asp | 120 | AGT | Ser | 125 | TTT | Phe | 130 | AAC | Asn | 135 | CAC | His | 140 | CGT | Arg | 145 | TAT | Tyr | 150 | AGC | Ser | 155 | ACT | Thr | 160 |     |     |     |
| 65 | TGG GCT GGC | Trp Ala Gly | 110 | AAG | Lys | 115 | GAA | Glu | 120 | CGG | Arg | 125 | TTG | Leu | 130 | AAA | Lys | 135 | TTC | Phe | 140 | TAT | Tyr | 145 | GCC | Ala | 150 | GGG | Gly | 155 | GAT | Asp | 160 | TAT | Tyr | 165 | ATT | Ile | 170 | GAT | Asp | 175 |     |     |     |
| 70 | CCA ACA TTG | Pro Thr Leu | 125 | CGA | Arg | 130 | TTG | Leu | 135 | AAT | Asn | 140 | AAG | Lys | 145 | ACC | Thr | 150 | GAG | Glu | 155 | ATA | Ile | 160 | TTT | Phe | 165 | ACC | Thr | 170 | GCA | Ala | 175 | TTT | Phe | 180 | GAA | Glu | 185 | CAA | Gln | 190 |     |     |     |

|    |                                                                      |
|----|----------------------------------------------------------------------|
|    | GGT ATT TCT CAA GGG AAA TTA AAA AGT GAA TTA GTC GAA TCT AAA TTA 483  |
|    | Gly Ile Ser Gln Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu      |
|    | 145 150 155                                                          |
| 5  | CGT GAT TAT CTA ATT AGT TAT GAC ACT TTA GCC ACC CTT GAT TAT ATT 531  |
|    | Arg Asp Tyr Leu Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile      |
|    | 160 165 170                                                          |
| 10 | ACT GCC TGC CAA GGC AAA GAT AAT AAA ACC ATC TTC TTT ATT GGC CGT 579  |
|    | Thr Ala Cys Gln Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg      |
|    | 175 180 185                                                          |
| 15 | ACA CAG AAT GCA CCC TAT GCA TTT TAT TGG CGA AAA TTA ACT TTA GTC 627  |
|    | Thr Gln Asn Ala Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val      |
|    | 190 195 200                                                          |
| 20 | ACT GAT GGC GGT AAG TTG AAA CCA GAT CAA TGG TCA GAG TGG CGA GCA 675  |
|    | Thr Asp Gly Gly Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala      |
|    | 205 210 215 220                                                      |
| 25 | ATT AAT GCC GGG ATT AGT GAG GCA TAT TCA GGG CAT GTC GAG CCT TTC 723  |
|    | Ile Asn Ala Gly Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe      |
|    | 225 230 235                                                          |
| 30 | TGG GAA AAT AAC AAG CTG CAC ATC CGT TGG TTT ACT ATC TCG AAA GAA 771  |
|    | Trp Glu Asn Asn Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu      |
|    | 240 245 250                                                          |
| 35 | GAT AAA ATA GAT TTT GTT TAT AAA AAC ATC TGG GTG ATG AGT ACC GAT 819  |
|    | Asp Lys Ile Asp Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp      |
|    | 255 260 265                                                          |
| 40 | TAT AGC TGG GCA TCA AAG AAA AAA ATC TTG GAA CTT TCT TTT ACT GAC 867  |
|    | Tyr Ser Trp Ala Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp      |
|    | 270 275 280                                                          |
| 45 | TAC AAT AGA GTT GGA GCA ACA GGA TCA TCA AGC CCG ACT GAA GTA GCT 915  |
|    | Tyr Asn Arg Val Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala      |
|    | 285 290 295 300                                                      |
| 50 | TCA CAA TAT GGT TCT GAT GCT CAG ATG AAT ATT TCT GAT GAT GGG ACT 963  |
|    | Ser Gln Tyr Gly Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr      |
|    | 305 310 315                                                          |
| 55 | GTA CTT ATT TTT CAG AAT GCC GGC GGA GCT ACT CCC AGT ACT GGA GTG 1011 |
|    | Val Leu Ile Phe Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val      |
|    | 320 325 330                                                          |
| 60 | ACG TTA TGT TAT GAC TCT GGC AAC GTG ATT AAG AAC CTA TCT AGT ACA 1059 |
|    | Thr Leu Cys Tyr Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr      |
|    | 335 340 345                                                          |
| 65 | GGA AGT GCA AAT TTA TCG TCA AAG GAT TAT GCC ACA ACT AAA TTA CGC 1107 |
|    | Gly Ser Ala Asn Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg      |
|    | 350 355 360                                                          |
| 70 | ATG TGT CAT GGA CAA AGT TAC AAT GAT AAT AAC TAC TGC AAT TTT ACA 1155 |
|    | Met Cys His Gly Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr      |
|    | 365 370 375 380                                                      |
| 75 | CTC TCT ATT AAT ACA ATA GAA TTC ACC TCC TAC GGC ACA TTC TCA TCA 1203 |
|    | Leu Ser Ile Asn Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser      |
|    | 385 390 395                                                          |
| 80 | GAT GGA AAA CAA TTT ACA CCA CCT TCT GGT TCT GCC ATT GAT TTA CAC 1251 |
|    | Asp Gly Lys Gln Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His      |

|    | 400                                                                                                                                                        | 405 | 410 |  |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-----|--|
| 5  | CTC CCT AAT TAT GTA GAT CTC AAC GCG CTA TTA GAT ATT AGC CTC GAT 1299<br>Leu Pro Asn Tyr Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp<br>415 420 425     |     |     |  |
| 10 | TCA CTA CTT AAT TAT GAC GTT CAG GGG CAG TTT GGC GGA TCT AAT CCG 1347<br>Ser Leu Leu Asn Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro<br>430 435 440     |     |     |  |
| 15 | GTT GAT AAT TTC AGT GGT CCC TAT GGT ATT TAT CTA TGG GAA ATC TTC 1395<br>Val Asp Asn Phe Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe<br>445 450 455 460 |     |     |  |
| 20 | TTC CAT ATT CCG TTC CTT GTT ACG GTC CGT ATG CAA ACC GAA CAA CGT 1443<br>Phe His Ile Pro Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg<br>465 470 475     |     |     |  |
| 25 | TAC GAA GAC GCG GAC ACT TGG TAC AAA TAT ATT TTC CGC AGC GCC GGT 1491<br>Tyr Glu Asp Ala Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly<br>480 485 490     |     |     |  |
| 30 | TAT CGC GAT GCT AAT GGC CAG CTC ATT ATG GAT GGC AGT AAA CCA CGT 1539<br>Tyr Arg Asp Ala Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg<br>495 500 505     |     |     |  |
| 35 | TAT TGG AAT GTG ATG CCA TTG CAA CTG GAT ACC GCA TGG GAT ACC ACA 1587<br>Tyr Trp Asn Val Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr<br>510 515 520     |     |     |  |
| 40 | CAG CCC GCC ACC ACT GAT CCA GAT GTG ATC GCT ATG GCG GAC CCG ATG 1635<br>Gln Pro Ala Thr Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met<br>525 530 535 540 |     |     |  |
| 45 | CAT TAC AAG CTG GCG ATA TTC CTG CAT ACC CTT GAT CTA TTG ATT GCC 1683<br>His Tyr Lys Leu Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala<br>545 550 555     |     |     |  |
| 50 | CGA GGC GAC AGC GCT TAC CGT CAA CTT GAA CGC GAT ACT CTA GTC GAA 1731<br>Arg Gly Asp Ser Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu<br>560 565 570     |     |     |  |
| 55 | GCC AAA ATG TAC TAC ATT CAG GCA CAA CAG CTA CTG GGA CCG CGC CCT 1779<br>Ala Lys Met Tyr Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro<br>575 580 585     |     |     |  |
| 60 | GAT ATC CAT ACC ACC AAT ACT TGG CCA AAT CCC ACC TTG AGT AAA GAA 1827<br>Asp Ile His Thr Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu<br>590 595 600     |     |     |  |
| 65 | GCT GGC GCT ATT GCC ACA CCG ACA TTC CTC AGT TCA CCG GAG GTG ATG 1875<br>Ala Gly Ala Ile Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met<br>605 610 615 620 |     |     |  |
| 70 | ACG TTC GCT GCC TGG CTA AGC GCA GGC GAT ACC GCA AAT ATT GGC GAC 1923<br>Thr Phe Ala Ala Trp Leu Ser Ala Gly Asp Thr Ala Asn Ile Gly Asp<br>625 630 635     |     |     |  |
| 75 | GGT GAT TTC TTG CCA CCG TAC AAC GAT GTA CTA CTC GGT TAC TGG GAT 1971<br>Gly Asp Phe Leu Pro Pro Tyr Asn Asp Val Leu Leu Gly Tyr Trp Asp<br>640 645 650     |     |     |  |
| 80 | AAA CTT GAG TTA CGC CTA TAC AAC CTG CGC CAC AAT CTG AGT CTG GAT 2019<br>Lys Leu Glu Leu Arg Leu Tyr Asn Leu Arg His Asn Leu Ser Leu Asp<br>655 660 665     |     |     |  |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | 3GT | CAA | CCG | CTA | AAT | CTG | CCA | CTG | TAT | GCC | ACG | CCG | GTA | GAC | CCG | AAA | 2057 |
|    | Gly | Gln | Pro | Leu | Asn | Leu | Pro | Leu | Tyr | Ala | Thr | Pro | Val | Asp | Pro | Lys |      |
|    | 570 |     |     |     |     |     | 675 |     |     |     |     | 680 |     |     |     |     |      |
| 5  | ACC | CTG | CAA | CGC | CAG | CAA | GCC | GGA | GGG | GAC | GGT | ACA | GGC | AGT | AGT | CCG | 2115 |
|    | Thr | Leu | Gln | Arg | Gln | Gln | Ala | Gly | Gly | Asp | Gly | Thr | Gly | Ser | Ser | Pro |      |
|    | 685 |     |     |     |     | 690 |     |     |     | 695 |     |     |     |     |     | 700 |      |
| 10 | GCT | GGT | GGT | CAA | GGC | AGT | GTT | CAG | GGC | TGG | CGC | TAT | CCG | TTA | TTG | GTA | 2163 |
|    | Ala | Gly | Gly | Gln | Gly | Ser | Val | Gln | Gly | Trp | Arg | Tyr | Pro | Leu | Leu | Val |      |
|    |     |     |     |     | 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |      |
| 15 | GAA | CGC | GCC | CGC | TCT | GCC | GTG | AGT | TTG | TTG | ACT | CAG | TTC | GGC | AAC | AGC | 2211 |
|    | Glu | Arg | Ala | Arg | Ser | Ala | Val | Ser | Leu | Leu | Thr | Gln | Phe | Gly | Asn | Ser |      |
|    |     |     |     | 720 |     |     |     |     | 725 |     |     |     |     | 730 |     |     |      |
| 20 | TTA | CAA | ACA | ACG | TTA | GAA | CAT | CAG | GAT | AAT | GAA | AAA | ATG | ACG | ATA | CTG | 2259 |
|    | Leu | Gln | Thr | Thr | Leu | Glu | His | Gln | Asp | Asn | Glu | Lys | Met | Thr | Ile | Leu |      |
|    |     |     | 735 |     |     |     | 740 |     |     |     |     |     | 745 |     |     |     |      |
|    | TTG | CAG | ACT | CAA | CAG | GAA | GCC | ATC | CTG | AAA | CAT | CAG | CAC | GAT | ATA | CAA | 2307 |
|    | Leu | Gln | Thr | Gln | Gln | Glu | Ala | Ile | Leu | Lys | His | Gln | His | Asp | Ile | Gln |      |
|    |     | 750 |     |     |     |     | 755 |     |     |     |     | 760 |     |     |     |     |      |
| 25 | CAA | AAT | AAT | CTA | AAA | GGA | TTA | CAA | CAC | AGC | CTG | ACC | GCA | TTA | CAG | GCT | 2355 |
|    | Gln | Asn | Asn | Leu | Lys | Gly | Leu | Gln | His | Ser | Leu | Thr | Ala | Leu | Gln | Ala |      |
|    | 765 |     |     |     |     | 770 |     |     |     | 775 |     |     |     |     |     | 780 |      |
| 30 | AGC | CGT | GAT | GGC | GAC | ACA | TTG | CGG | CAA | AAA | CAT | TAC | AGC | GAC | CTG | ATT | 2403 |
|    | Ser | Arg | Asp | Gly | Asp | Thr | Leu | Arg | Gln | Lys | His | Tyr | Ser | Asp | Leu | Ile |      |
|    |     |     |     |     | 785 |     |     |     |     | 790 |     |     |     |     | 795 |     |      |
| 35 | AAC | GGT | GGT | CTA | TCT | GCG | GCA | GAA | ATC | GCC | GGT | CTG | ACA | CTA | CGC | AGC | 2451 |
|    | Asn | Gly | Gly | Leu | Ser | Ala | Ala | Glu | Ile | Ala | Gly | Leu | Thr | Leu | Arg | Ser |      |
|    |     |     |     | 800 |     |     |     |     | 805 |     |     |     |     | 810 |     |     |      |
|    | ACC | GCC | ATG | ATT | ACC | AAT | GGC | GTT | GCA | ACG | GGA | TTG | CTG | ATT | GCC | GGC | 2499 |
|    | Thr | Ala | Met | Ile | Thr | Asn | Gly | Val | Ala | Thr | Gly | Leu | Leu | Ile | Ala | Gly |      |
| 40 |     |     | 815 |     |     |     |     | 820 |     |     |     |     | 825 |     |     |     |      |
|    | GGA | ATC | GCC | AAC | GCG | GTA | CCT | AAC | GTC | TTC | GGG | CTG | GCT | AAC | GGT | GGA | 2547 |
|    | Gly | Ile | Ala | Asn | Ala | Val | Pro | Asn | Val | Phe | Gly | Leu | Ala | Asn | Gly | Gly |      |
|    |     | 830 |     |     |     |     | 835 |     |     |     |     | 840 |     |     |     |     |      |
| 45 | TCG | GAA | TGG | GGA | GCG | CCA | TTA | ATT | GGC | TCC | GGG | CAA | GCA | ACC | CAA | GTT | 2595 |
|    | Ser | Glu | Trp | Gly | Ala | Pro | Leu | Ile | Gly | Ser | Gly | Gln | Ala | Thr | Gln | Val |      |
|    | 845 |     |     |     |     | 850 |     |     |     | 855 |     |     |     |     |     | 860 |      |
| 50 | GGC | GCC | GGC | ATC | CAG | GAT | CAG | AGC | GCG | GGC | ATT | TCA | GAA | GTG | ACA | GCA | 2643 |
|    | Gly | Ala | Gly | Ile | Gln | Asp | Gln | Ser | Ala | Gly | Ile | Ser | Glu | Val | Thr | Ala |      |
|    |     |     |     |     | 865 |     |     |     |     | 870 |     |     |     |     | 875 |     |      |
| 55 | GGC | TAT | CAG | CGT | CGT | CAG | GAA | GAA | TGG | GCA | TTG | CAA | CGG | GAT | ATT | GCT | 2691 |
|    | Gly | Tyr | Gln | Arg | Arg | Gln | Glu | Glu | Trp | Ala | Leu | Gln | Arg | Asp | Ile | Ala |      |
|    |     |     |     | 880 |     |     |     |     | 885 |     |     |     |     | 890 |     |     |      |
|    | GAT | AAC | GAA | ATA | ACC | CAA | CTG | GAT | GCC | CAG | ATA | CAA | AGC | CTG | CAA | GAG | 2739 |
|    | Asp | Asn | Glu | Ile | Thr | Gln | Leu | Asp | Ala | Gln | Ile | Gln | Ser | Leu | Gln | Glu |      |
| 60 |     |     | 895 |     |     |     |     | 900 |     |     |     |     | 905 |     |     |     |      |
|    | CAA | ATC | ACG | ATG | GCA | CAA | AAA | CAG | ATC | ACG | CTC | TCT | GAA | ACC | GAA | CAA | 2787 |
|    | Gln | Ile | Thr | Met | Ala | Gln | Lys | Gln | Ile | Thr | Leu | Ser | Glu | Thr | Glu | Gln |      |
|    |     | 910 |     |     |     |     | 915 |     |     |     |     | 920 |     |     |     |     |      |
| 65 | GCG | AAT | GCC | CAA | GCG | ATT | TAT | GAC | CTG | CAA | ACC | ACT | CGT | TTT | ACC | GGG | 2835 |
|    | Ala | Asn | Ala | Gln | Ala | Ile | Tyr | Asp | Leu | Gln | Thr | Thr | Arg | Phe | Thr | Gly |      |

|    | 925                                                                  |  | 930                                                                  |  | 935 |  | 940 |  |
|----|----------------------------------------------------------------------|--|----------------------------------------------------------------------|--|-----|--|-----|--|
| 5  | CAG GCA CTG TAT AAC TGG ATG GCC GGT CGT CTC TCC GCG CTC TAT TAC 2333 |  | Gln Ala Leu Tyr Asn Trp Met Ala Gly Arg Leu Ser Ala Leu Tyr Tyr 955  |  |     |  |     |  |
| 10 | CAA ATG TAT GAT TCC ACT CTG CCA ATC TGT CTC CAG CCA AAA GCC GCA 2931 |  | Gln Met Tyr Asp Ser Thr Leu Pro Ile Cys Leu Gln Pro Lys Ala Ala 970  |  |     |  |     |  |
| 15 | TTA GTA CAG GAA TTA GGC GAG AAA GAG AGC GAC AGT CTT TTC CAG GTT 2979 |  | Leu Val Gln Glu Leu Gly Glu Lys Glu Ser Asp Ser Leu Phe Gln Val 985  |  |     |  |     |  |
| 20 | CCG GTG TGG AAT GAT CTG TGG CAA GGG CTG TTA GCA GGA GAA GGT TTA 3027 |  | Pro Val Trp Asn Asp Leu Trp Gln Gly Leu Leu Ala Gly Glu Gly Leu 1000 |  |     |  |     |  |
| 25 | AGT TCA GAG CTA CAG AAA CTG GAT GCC ATC TGG CTT GCA CGT GGT GGT 3075 |  | Ser Ser Glu Leu Gln Lys Leu Asp Ala Ile Trp Leu Ala Arg Gly Gly 1020 |  |     |  |     |  |
| 30 | ATT GGG CTA GAA GCC ATC CGC ACC GTG TCG CTG GAT ACC CTG TTT GGC 3123 |  | Ile Gly Leu Glu Ala Ile Arg Thr Val Ser Leu Asp Thr Leu Phe Gly 1035 |  |     |  |     |  |
| 35 | ACA GGG ACG TTA AGT GAA AAT ATC AAT AAA GTG CTT AAC GGG GAA ACG 3171 |  | Thr Gly Thr Leu Ser Glu Asn Ile Asn Lys Val Leu Asn Gly Glu Thr 1050 |  |     |  |     |  |
| 40 | GTA TCT CCA TCC GGT GGC GTC ACT CTG GCG CTG ACA GGG GAT ATC TTC 3219 |  | Val Ser Pro Ser Gly Gly Val Thr Leu Ala Leu Thr Gly Asp Ile Phe 1065 |  |     |  |     |  |
| 45 | CAA GCA ACA CTG GAT TTG AGT CAG CTA GGT TTG GAT AAC TCT TAC AAC 3267 |  | Gln Ala Thr Leu Asp Leu Ser Gln Leu Gly Leu Asp Asn Ser Tyr Asn 1080 |  |     |  |     |  |
| 50 | TTG GGT AAC GAG AAG AAA CGT CGT ATT AAA CGT ATC GCC GTC ACC CTG 3315 |  | Leu Gly Asn Glu Lys Lys Arg Arg Ile Lys Arg Ile Ala Val Thr Leu 1100 |  |     |  |     |  |
| 55 | CCA ACA CTT CTG GGG CCA TAT CAA GAT CTT GAA GCC ACA CTG GTA ATG 3363 |  | Pro Thr Leu Leu Gly Pro Tyr Gln Asp Leu Glu Ala Thr Leu Val Met 1115 |  |     |  |     |  |
| 60 | GGT GCG GAA ATC GCC GCC TTA TCA CAC GGT GTG AAT GAC GGA GGC CGG 3411 |  | Gly Ala Glu Ile Ala Ala Leu Ser His Gly Val Asn Asp Gly Gly Arg 1130 |  |     |  |     |  |
| 65 | TTT GTT ACC GAC TTT AAC GAC AGC CGT TTT CTG CCT TTT GAA GGT CGA 3459 |  | Phe Val Thr Asp Phe Asn Asp Ser Arg Phe Leu Pro Phe Glu Gly Arg 1145 |  |     |  |     |  |
|    | GAT GCA ACA ACC GGC ACA CTG GAG CTC AAT ATT TTC CAT GCG GGT AAA 3507 |  | Asp Ala Thr Thr Gly Thr Leu Glu Leu Asn Ile Phe His Ala Gly Lys 1160 |  |     |  |     |  |
|    | GAG GGA ACG CAA CAC GAG TTG GTC GCG AAT CTG AGT GAC ATC ATT GTG 3555 |  | Glu Gly Thr Gln His Glu Leu Val Ala Asn Leu Ser Asp Ile Ile Val 1180 |  |     |  |     |  |
|    | CAT CTG AAT TAC ATC ATT CGA GAC GCG TAA ATTTCTTTTC TTGTCGATT 3605    |  | His Leu Asn Tyr Ile Ile Arg Asp Ala 1190                             |  |     |  |     |  |

ACAGGTCCTT ATCAGGGGCC TGTATTAAAG GAGTACTTTA TGCAGGATTC ACCAGAAGTA 3555  
 TCGATTACAA CGCTGTCACT TCCCAAAGGT GGCGGTGCTA TCAATGGCAT GGGAGAAGCA 3725  
 5 CTGAATGCTG CCGGCCCTGA TCGAATGGCC TCCCTATCTC TGCCATTACC CCTTTCGACC 3785  
 GGCAGAGGGA CGGCTCCTGG ATTATCGCTG ATTTACAGCA ACAGTGCAGG TAATGGGCCT 3845  
 10 TTCGGCATCG GCTGGCAATG CGGTGTTATG TCCATTAGCC GACGCACCCA ACATGGCATT 3905  
 CCACAATACG GTAATGACGA CACGTTCCCTA TCCCCACAAG GCGAGGTCAT GAATATCGCC 3965  
 CTGAATGACC AAGGGCAACC TGATATCCGT CAAGACGTTA AAACGCTGCA AGGCGTTACC 4025  
 15 TTGCCAATTT CCTATACCGT GACCCGCTAT CAAGCCCGCC AGATCCTGGA TTTCAGTAAA 4085  
 ATCGAATACT GGCAACCTGC CTCCGGTCAA GAAGGACGCG CTTTCTGGCT GATATCGACA 4145  
 CCGGACGGGC ATCTACACAT CTTAGGGAAA ACCGCGCAGG CTTGTCTGGC AAATCCGCAA 4205  
 20 AATGACCAAC AAATCGCCCA GTGGTTGCTG GAAGAACTG TGACGCCAGC CGGTGAACAT 4265  
 GTCAGCTATC AATATCGAGC CGAAGATGAA GCCCATTTGTG ACGACAATGA AAAAACCGCT 4325  
 25 CATCCCAATG TTACCGCACA GCGCTATCTG GTACAGGTGA ACTACAGGCA ACATCAAACC 4385  
 ACAAGCCAGC CTGTTCTGAC TGGATAACGC ACCTCCCGCA CCGGAAGAGT GGCTGTTTCA 4445  
 TCTGGTCTTT GACCACGGTG AGCGCGTACC TCACTTCATA CCGTGCCAAC ATGGGATGCA 4505  
 30 GGTACAGCGC AATGGTCTGT ACGCCCGGAT ATCTTCTCTC GCTATGAATA TGGTTTTGAA 4565  
 GTGCGTACTC GCCGCTTATG TCAACAAGTG CTGATGTTTC ACCGCACCGC GCTCATGGCC 4625  
 35 GGAGAAGCCA GTACCAATGA CGCCCCGGA CTGGTTGGAC GCTTAATACT GGAATATGAC 4685  
 AAAAACGCCA GCGTCACCAC GTTGATTACC ATCCGTCAAT TAAGCCATGA ATCGGACGGG 4745  
 AGGCCAGTCA CCCAGCCACC ACTAGAACTA GCCTGGCAAC GGTTTGATCT GGAGAAAATC 4805  
 40 CCGACATGGC AACGCTTTGA CGCACTAGAT AATTTTAACT CGCAGCAACG TTATCAACTG 4865  
 GTTGATCTGC GGGGAGAAGG GTTGCCAGGT ATGCTGTATC AAGATCGAGG CGCTTGGTGG 4925  
 45 TATAAAGCTC CGCAACGTCA GGAAGACGGA GACAGCAATG CCGTCACTTA CGACAAAATC 4985  
 GCCCCACTGC CTACCCTACC CAATTTGCAG GATAATGCCT CATTGATGGA TATCAACGGA 5045  
 GACGGCCAAC TGGATTGGGT TGTTACCGCC TCCGGTATTC GCGGATACCA TAGTCAGCAA 5105  
 50 CCCGATGGAA AGTGGACGCA CTTTACGCCA ATCAATGCCT TGCCCGTGGG ATATTTTCAT 5165  
 CCAAGCATCC AGTTCGCTGA CCTTACCGGG GCAGGCTTAT CTGATTTAGT GTTGATCGGG 5225  
 55 CCGAAAAGCG TGCGTCTATA TGCCAACCAG CGAAACGGCT GGCGTAAAGG AGAAGATGTC 5285  
 CCCCCAATCA CAGGTATCAC CCTGCCTGTC ACAGGGACCG ATGCCCAGCA ACTGGTGGCT 5345  
 TTCAGTGATA TGCTCGGTTC CGGTCAACAA CATCTGGTGG AAATCAAGGG TAATCGCGTC 5405  
 60 ACCTGTTGGC CGAATCTAGG GCATGGCCGT TTCGGTCAAC CACTAACTCT GTCAGGATTT 5465  
 AGCCAGCCCG AAAATAGCTT CAATCCCGAA CGGCTGTTTC TGGCGGATAT CGACGGCTCC 5525  
 65 GGCACCACCG ACCTTATCTA TGCGCAATCC GGCTCTTTGC TCATTTATCT CAACCAAAGT 5585



5 GGTAAATCAGT TTGATGCCCC GTTGACATTA GCGTTGCCAG AAGGCGTACA ATTTGACAAC 5645  
 ACTTGCCAAC TTCAAGTCGC CGATATTCAG GGATTAGGGA TAGCCAGCTT GATTCTGACT 5705  
 10 GTGCCACATA TCGCGCCACA TCACTGGCGT TGTGACCTGT CACTGACCAA ACCCTGGTTG 5765  
 TTGAATGTAA TGAACAATAA CCGGGGCGCA CATCACACGC TACATTATCG TAGTTCCGCG 5825  
 CAATTCTGGT TGGATGAAAA ATTACAGCTC ACCAAAGCAG GCAAATCTCC GGCTTGTTAT 5885  
 15 CTGCCGTTTC CAATGCATTT GCTATGGTAT ACCGAAATTC AGGATGAAAT CAGCGGCAAC 5945  
 CGGCTCACCA GTGAAGTCAA CTACAGCCAC GCGCTCTGGG ATGGTAAAGA GCGGGAATTC 6005

15

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1190 amino acids  
 20 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp  
 1 5 10 15  
 30 Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro Ala Asp Leu Lys  
 20 25 30  
 35 Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr Leu Leu Leu Asp  
 35 40 45  
 Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile  
 50 55 60  
 40 Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu Gly Tyr Asp Gly  
 65 70 75 80  
 Thr Leu Ala Asp Ser Ala Lys Pro Tyr Phe Ala Asp Glu Gln Phe Leu  
 85 90 95  
 45 Tyr Asn Trp Asp Ser Phe Asn His Arg Tyr Ser Thr Trp Ala Gly Lys  
 100 105 110  
 50 Glu Arg Leu Lys Phe Tyr Ala Gly Asp Tyr Ile Asp Pro Thr Leu Arg  
 115 120 125  
 Leu Asn Lys Thr Glu Ile Phe Thr Ala Phe Glu Gln Gly Ile Ser Gln  
 130 135 140  
 55 Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu Arg Asp Tyr Leu  
 145 150 155 160  
 Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile Thr Ala Cys Gln  
 165 170 175  
 60 Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg Thr Gln Asn Ala  
 180 185 190  
 65 Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val Thr Asp Gly Gly  
 195 200 205

Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala Ile Asn Ala Gly  
 210 215 220  
 5 Ile Ser Glu Ala Tyr S r Gly His Val Glu Pro Phe Trp Glu Asn Asn  
 225 230 235 240  
 Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu Asp Lys Ile Asp  
 245 250 255  
 10 Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp Tyr Ser Trp Ala  
 260 265 270  
 Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp Tyr Asn Arg Val  
 275 280 285  
 Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala Ser Gln Tyr Gly  
 290 295 300  
 20 Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr Val Leu Ile Phe  
 305 310 315 320  
 Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val Thr Leu Cys Tyr  
 325 330 335  
 25 Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr Gly Ser Ala Asn  
 340 345 350  
 Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg Met Cys His Gly  
 355 360 365  
 Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr Leu Ser Ile Asn  
 370 375 380  
 35 Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln  
 385 390 395 400  
 Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr  
 405 410 415  
 40 Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn  
 420 425 430  
 Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe  
 435 440 445  
 Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro  
 450 455 460  
 50 Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala  
 465 470 475 480  
 Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala  
 485 490 495  
 55 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val  
 500 505 510  
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr  
 515 520 525  
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu  
 530 535 540  
 65 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser  
 545 550 555 560

Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr  
565 570 575

5 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr  
580 585 590

Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile  
595 600 605

10 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala  
610 615 620

15 Trp Leu Ser Ala Gly Asp Thr Ala Asn Ile Gly Asp Gly Asp Phe Leu  
625 630 635 640

Pro Pro Tyr Asn Asp Val Leu Leu Gly Tyr Trp Asp Lys Leu Glu Leu  
645 650 655

20 Arg Leu Tyr Asn Leu Arg His Asn Leu Ser Leu Asp Gly Gln Pro Leu  
660 665 670

Asn Leu Pro Leu Tyr Ala Thr Pro Val Asp Pro Lys Thr Leu Gln Arg  
675 680 685

25 Gln Gln Ala Gly Gly Asp Gly Thr Gly Ser Ser Pro Ala Gly Gly Gln  
690 695 700

30 Gly Ser Val Gln Gly Trp Arg Tyr Pro Leu Leu Val Glu Arg Ala Arg  
705 710 715 720

Ser Ala Val Ser Leu Leu Thr Gln Phe Gly Asn Ser Leu Gln Thr Thr  
725 730 735

35 Leu Glu His Gln Asp Asn Glu Lys Met Thr Ile Leu Leu Gln Thr Gln  
740 745 750

Gln Glu Ala Ile Leu Lys His Gln His Asp Ile Gln Gln Asn Asn Leu  
755 760 765

40 Lys Gly Leu Gln His Ser Leu Thr Ala Leu Gln Ala Ser Arg Asp Gly  
770 775 780

45 Asp Thr Leu Arg Gln Lys His Tyr Ser Asp Leu Ile Asn Gly Gly Leu  
785 790 795 800

Ser Ala Ala Glu Ile Ala Gly Leu Thr Leu Arg Ser Thr Ala Met Ile  
805 810 815

50 Thr Asn Gly Val Ala Thr Gly Leu Leu Ile Ala Gly Gly Ile Ala Asn  
820 825 830

Ala Val Pro Asn Val Phe Gly Leu Ala Asn Gly Gly Ser Glu Trp Gly  
835 840 845

55 Ala Pro Leu Ile Gly Ser Gly Gln Ala Thr Gln Val Gly Ala Gly Ile  
850 855 860

60 Gln Asp Gln Ser Ala Gly Ile Ser Glu Val Thr Ala Gly Tyr Gln Arg  
865 870 875 880

Arg Gln Glu Glu Trp Ala Leu Gln Arg Asp Ile Ala Asp Asn Glu Ile  
885 890 895

65 Thr Gln Leu Asp Ala Gln Ile Gln Ser Leu Gln Glu Gln Ile Thr Met  
900 905 910

Ala Gln Lys Gln Ile Thr Leu Ser Glu Thr Glu Gln Ala Asn Ala Gln  
 915 920 925  
 5 Ala Ile Tyr Asp Leu Gln Thr Thr Arg Phe Thr Gly Gln Ala Leu Tyr  
 930 935 940  
 Asn Trp Met Ala Gly Arg Leu Ser Ala Leu Tyr Tyr Gln Met Tyr Asp  
 945 950 955 960  
 10 Ser Thr Leu Pro Ile Cys Leu Gln Pro Lys Ala Ala Leu Val Gln Glu  
 965 970 975  
 15 Leu Gly Glu Lys Glu Ser Asp Ser Leu Phe Gln Val Pro Val Trp Asn  
 980 985 990  
 Asp Leu Trp Gln Gly Leu Leu Ala Gly Glu Gly Leu Ser Ser Glu Leu  
 995 1000 1005  
 20 Gln Lys Leu Asp Ala Ile Trp Leu Ala Arg Gly Gly Ile Gly Leu Glu  
 1010 1015 1020  
 Ala Ile Arg Thr Val Ser Leu Asp Thr Leu Phe Gly Thr Gly Thr Leu  
 1025 1030 1035 1040  
 25 Ser Glu Asn Ile Asn Lys Val Leu Asn Gly Glu Thr Val Ser Pro Ser  
 1045 1050 1055  
 30 Gly Gly Val Thr Leu Ala Leu Thr Gly Asp Ile Phe Gln Ala Thr Leu  
 1060 1065 1070  
 Asp Leu Ser Gln Leu Gly Leu Asp Asn Ser Tyr Asn Leu Gly Asn Glu  
 1075 1080 1085  
 35 Lys Lys Arg Arg Ile Lys Arg Ile Ala Val Thr Leu Pro Thr Leu Leu  
 1090 1095 1100  
 Gly Pro Tyr Gln Asp Leu Glu Ala Thr Leu Val Met Gly Ala Glu Ile  
 1105 1110 1115 1120  
 40 Ala Ala Leu Ser His Gly Val Asn Asp Gly Gly Arg Phe Val Thr Asp  
 1125 1130 1135  
 45 Phe Asn Asp Ser Arg Phe Leu Pro Phe Glu Gly Arg Asp Ala Thr Thr  
 1140 1145 1150  
 Gly Thr Leu Glu Leu Asn Ile Phe His Ala Gly Lys Glu Gly Thr Gln  
 1155 1160 1165  
 50 His Glu Leu Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr  
 1170 1175 1180  
 Ile Ile Arg Asp Ala \*  
 1185 1190  
 55

## (2) INFORMATION FOR SEQ ID NO:27:

- 60 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1881 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 65 (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- 5 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..1881  
 (D) OTHER INFORMATION: /product= "P8"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

|    |                                                                 |     |
|----|-----------------------------------------------------------------|-----|
| 10 | ATG TCT GAA TCT TTA TTT ACA CAA ACG TTG AAA GAA GCG CGC CGT GAT | 48  |
|    | Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp |     |
|    | 1 5 10 15                                                       |     |
| 15 | GCA TTG GTT GCT CAT TAT ATT GCT ACT CAG GTG CCC GCA GAT TTA AAA | 96  |
|    | Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro Ala Asp Leu Lys |     |
|    | 20 25 30                                                        |     |
| 20 | GAG AGT ATC CAG ACC GCG GAT GAT CTG TAC GAA TAT CTG TTG CTG GAT | 144 |
|    | Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr Leu Leu Asp     |     |
|    | 35 40 45                                                        |     |
| 25 | ACC AAA ATT AGC GAT CTG GTT ACT ACT TCA CCG CTG TCC GAA GCG ATT | 192 |
|    | Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile |     |
|    | 50 55 60                                                        |     |
| 30 | GGC AGT CTG CAA TTG TTT ATT CAT CGT GCG ATA GAG GGC TAT GAC GGC | 240 |
|    | Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu Gly Tyr Asp Gly |     |
|    | 65 70 75 80                                                     |     |
| 35 | ACG CTG GCA GAC TCA GCA AAA CCC TAT TTT GCC GAT GAA CAG TTT TTA | 288 |
|    | Thr Leu Ala Asp Ser Ala Lys Pro Tyr Phe Ala Asp Glu Gln Phe Leu |     |
|    | 85 90 95                                                        |     |
| 40 | TAT AAC TGG GAT AGT TTT AAC CAC CGT TAT AGC ACT TGG GCT GGC AAG | 336 |
|    | Tyr Asn Trp Asp Ser Phe Asn His Arg Tyr Ser Thr Trp Ala Gly Lys |     |
|    | 100 105 110                                                     |     |
| 45 | GAA CGG TTG AAA TTC TAT GCC GGG GAT TAT ATT GAT CCA ACA TTG CGA | 384 |
|    | Glu Arg Leu Lys Phe Tyr Ala Gly Asp Tyr Ile Asp Pro Thr Leu Arg |     |
|    | 115 120 125                                                     |     |
| 50 | TTG AAT AAG ACC GAG ATA TTT ACC GCA TTT GAA CAA GGT ATT TCT CAA | 432 |
|    | Leu Asn Lys Thr Glu Ile Phe Thr Ala Phe Glu Gln Gly Ile Ser Gln |     |
|    | 130 135 140                                                     |     |
| 55 | GGG AAA TTA AAA AGT GAA TTA GTC GAA TCT AAA TTA CGT GAT TAT CTA | 480 |
|    | Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu Arg Asp Tyr Leu |     |
|    | 145 150 155 160                                                 |     |
| 60 | ATT AGT TAT GAC ACT TTA GCC ACC CTT GAT TAT ATT ACT GCC TGC CAA | 528 |
|    | Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile Thr Ala Cys Gln |     |
|    | 165 170 175                                                     |     |
| 65 | GGC AAA GAT AAT AAA ACC ATC TTC TTT ATT GGC CGT ACA CAG AAT GCA | 576 |
|    | Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg Thr Gln Asn Ala |     |
|    | 180 185 190                                                     |     |
| 70 | CCC TAT GCA TTT TAT TGG CGA AAA TTA ACT TTA GTC ACT GAT GGC GGT | 624 |
|    | Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val Thr Asp Gly Gly |     |
|    | 195 200 205                                                     |     |
| 75 | AAG TTG AAA CCA GAT CAA TGG TCA GAG TGG CGA GCA ATT AAT GCC GGG | 672 |
|    | Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala Ile Asn Ala Gly |     |
|    | 210 215 220                                                     |     |

|    |                                                                                                                                                            |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5  | ATT AGT GAG GCA TAT TCA GGG CAT CTC GAG CCT TTC TGG GAA AAT AAC 700<br>Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe Trp Glu Asn Asn<br>225 230 235 240  |
| 10 | AAG CTG CAC ATC CGT TGG TTT ACT ATC TCG AAA GAA GAT AAA ATA GAT 768<br>Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu Asp Lys Ile Asp<br>245 250 255      |
| 15 | TTT GTT TAT AAA AAC ATC TGG GTG ATG AGT AGC GAT TAT AGC TGG GCA 816<br>Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp Tyr Ser Trp Ala<br>260 265 270      |
| 20 | TCA AAG AAA AAA ATC TTG GAA CTT TCT TTT ACT GAC TAC AAT AGA GTT 864<br>Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp Tyr Asn Arg Val<br>275 280 285      |
| 25 | GGA GCA ACA GGA TCA TCA AGC CCG ACT GAA GTA GCT TCA CAA TAT GGT 912<br>Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala Ser Gln Tyr Gly<br>290 295 300      |
| 30 | TCT GAT GCT CAG ATG AAT ATT TCT GAT GAT GGG ACT GTA CTT ATT TTT 960<br>Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr Val Leu Ile Phe<br>305 310 315 320  |
| 35 | CAG AAT GCC GGC GGA GCT ACT CCC AGT ACT GGA GTG ACG TTA TGT TAT 1008<br>Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val Thr Leu Cys Tyr<br>325 330 335     |
| 40 | GAC TCT GGC AAC GTG ATT AAG AAC CTA TCT AGT ACA GGA AGT GCA AAT 1056<br>Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr Gly Ser Ala Asn<br>340 345 350     |
| 45 | TTA TCG TCA AAG GAT TAT GCC ACA ACT AAA TTA CGC ATG TGT CAT GGA 1104<br>Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg Met Cys His Gly<br>355 360 365     |
| 50 | CAA AGT TAC AAT GAT AAT AAC TAC TGC AAT TTT ACA CTC TCT ATT AAT 1152<br>Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr Leu Ser Ile Asn<br>370 375 380     |
| 55 | ACA ATA GAA TTC ACC TCC TAC GGC ACA TTC TCA TCA GAT GGA AAA CAA 1200<br>Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln<br>385 390 395 400 |
| 60 | TTT ACA CCA CCT TCT GGT TCT GCC ATT GAT TTA CAC CTC CCT AAT TAT 1248<br>Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr<br>405 410 415     |
| 65 | GTA GAT CTC AAC GCG CTA TTA GAT ATT AGC CTC GAT TCA CTA CTT AAT 1296<br>Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn<br>420 425 430     |
| 70 | TAT GAC GTT CAG GGG CAG TTT GGC GGA TCT AAT CCG GTT GAT AAT TTC 1344<br>Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe<br>435 440 445     |
| 75 | AGT GGT CCC TAT GGT ATT TAT CTA TGG GAA ATC TTC TTC CAT ATT CCG 1392<br>Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro<br>450 455 460     |
| 80 | TTC CTT GTT ACG GTC CGT ATG CAA ACC GAA CAA CGT TAC GAA GAC GCG 1440<br>Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala<br>465 470 475 480 |
| 85 | GAC ACT TGG TAC AAA TAT ATT TTC CGC AGC GCC GGT TAT CGC GAT GCT 1488                                                                                       |

Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala  
 485 490 495  
 5 AAT GGC CAG CTC ATT ATG GAT GGC AGT AAA CCA CGT TAT TGG AAT GTG 1535  
 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val  
 500 505 510  
 10 ATG CCA TTG CAA CTG GAT ACC GCA TGG GAT ACC ACA CAG CCC GCC ACC 1534  
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr  
 515 520 525  
 15 ACT GAT CCA GAT GTG ATC GCT ATG GCG GAC CCG ATG CAT TAC AAG CTG 1632  
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu  
 530 535 540  
 GCG ATA TTC CTG CAT ACC CTT GAT CTA TTG ATT GCC CGA GGC GAC AGC 1680  
 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser  
 545 550 555 560  
 20 GCT TAC CGT CAA CTT GAA CGC GAT ACT CTA GTC GAA GCC AAA ATG TAC 1728  
 Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr  
 565 570 575  
 25 TAC ATT CAG GCA CAA CAG CTA CTG GGA CCG CGC CCT GAT ATC CAT ACC 1776  
 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr  
 580 585 590  
 30 ACC AAT ACT TGG CCA AAT CCC ACC TTG AGT AAA GAA GCT GGC GCT ATT 1824  
 Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile  
 595 600 605  
 GCC ACA CCG ACA TTC CTC AGT TCA CCG GAG GTG ATG ACG TTC GCT GCC 1872  
 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala  
 610 615 620  
 35 TGG CTA AGC 1881  
 Trp Leu Ser  
 625

40

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 627 amino acids  
 45 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp  
 1 5 10 15  
 55 Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro Ala Asp Leu Lys  
 20 25 30  
 60 Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr Leu Leu Leu Asp  
 35 40 45  
 Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile  
 50 55 60  
 65 Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu Gly Tyr Asp Gly

|    | 65  |     | 70  |     | 75  |     | 80  |
|----|-----|-----|-----|-----|-----|-----|-----|
|    | Thr | Leu | Ala | Asp | Ser | Ala | Lys |
|    |     |     |     | 85  |     |     | 90  |
|    |     |     |     |     |     |     | 95  |
| 5  | Tyr | Asn | Trp | Asp | Ser | Phe | Asn |
|    |     |     | 100 |     |     |     | 105 |
|    |     |     |     |     |     |     | 110 |
|    | Glu | Arg | Leu | Lys | Phe | Tyr | Ala |
|    |     |     | 115 |     |     |     | 120 |
| 10 |     |     |     |     |     |     | 125 |
|    | Leu | Asn | Lys | Thr | Glu | Ile | Phe |
|    |     |     | 130 |     |     |     | 135 |
|    |     |     |     |     |     |     | 140 |
|    | Gly | Lys | Leu | Lys | Ser | Glu | Leu |
|    |     |     | 145 |     |     |     | 150 |
| 15 |     |     |     |     |     |     | 155 |
|    | Ile | Ser | Tyr | Asp | Thr | Leu | Ala |
|    |     |     |     | 165 |     |     | 170 |
|    |     |     |     |     |     |     | 175 |
| 20 | Gly | Lys | Asp | Asn | Lys | Thr | Ile |
|    |     |     |     | 180 |     |     | 185 |
|    |     |     |     |     |     |     | 190 |
|    | Pro | Tyr | Ala | Phe | Tyr | Trp | Arg |
|    |     |     | 195 |     |     |     | 200 |
| 25 |     |     |     |     |     |     | 205 |
|    | Lys | Leu | Lys | Pro | Asp | Gln | Trp |
|    |     |     | 210 |     |     |     | 215 |
|    |     |     |     |     |     |     | 220 |
|    | Ile | Ser | Glu | Ala | Tyr | Ser | Gly |
|    |     |     |     |     |     |     | 230 |
| 30 |     |     |     |     |     |     | 235 |
|    |     |     |     |     |     |     | 240 |
|    | Lys | Leu | His | Ile | Arg | Trp | Phe |
|    |     |     |     | 245 |     |     | 250 |
|    |     |     |     |     |     |     | 255 |
| 35 | Phe | Val | Tyr | Lys | Asn | Ile | Trp |
|    |     |     |     | 260 |     |     | 265 |
|    |     |     |     |     |     |     | 270 |
|    | Ser | Lys | Lys | Lys | Ile | Leu | Glu |
|    |     |     |     | 275 |     |     | 280 |
| 40 |     |     |     |     |     |     | 285 |
|    | Gly | Ala | Thr | Gly | Ser | Ser | Ser |
|    |     |     |     | 290 |     |     | 295 |
|    |     |     |     |     |     |     | 300 |
|    | Ser | Asp | Ala | Gln | Met | Asn | Ile |
|    |     |     |     |     |     |     | 310 |
| 45 |     |     |     |     |     |     | 315 |
|    |     |     |     |     |     |     | 320 |
|    | Gln | Asn | Ala | Gly | Gly | Ala | Thr |
|    |     |     |     | 325 |     |     | 330 |
|    |     |     |     |     |     |     | 335 |
| 50 | Asp | Ser | Gly | Asn | Val | Ile | Lys |
|    |     |     |     | 340 |     |     | 345 |
|    |     |     |     |     |     |     | 350 |
|    | Leu | Ser | Ser | Lys | Asp | Tyr | Ala |
|    |     |     |     | 355 |     |     | 360 |
| 55 |     |     |     |     |     |     | 365 |
|    | Gln | Ser | Tyr | Asn | Asp | Asn | Asn |
|    |     |     |     | 370 |     |     | 375 |
|    |     |     |     |     |     |     | 380 |
|    | Thr | Ile | Glu | Phe | Thr | Ser | Tyr |
|    |     |     |     |     |     |     | 390 |
| 60 |     |     |     |     |     |     | 395 |
|    |     |     |     |     |     |     | 400 |
|    | Phe | Thr | Pro | Pro | Ser | Gly | Ser |
|    |     |     |     |     |     |     | 405 |
|    |     |     |     |     |     |     | 410 |
|    |     |     |     |     |     |     | 415 |
| 65 | Val | Asp | Leu | Asn | Ala | Leu | Leu |
|    |     |     |     |     |     |     | Asp |
|    |     |     |     |     |     |     | Ile |
|    |     |     |     |     |     |     | Ser |
|    |     |     |     |     |     |     | Leu |
|    |     |     |     |     |     |     | Asp |
|    |     |     |     |     |     |     | Ser |
|    |     |     |     |     |     |     | Leu |
|    |     |     |     |     |     |     | Leu |
|    |     |     |     |     |     |     | Asn |



420 425 430  
 Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe  
 435 440 445  
 5 Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro  
 450 455 460  
 10 Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala  
 465 470 475 480  
 Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala  
 485 490 495  
 15 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val  
 500 505 510  
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr  
 515 520 525  
 20 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu  
 530 535 540  
 25 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser  
 545 550 555 560  
 Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr  
 565 570 575  
 30 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr  
 580 585 590  
 Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile  
 595 600 605  
 35 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala  
 610 615 620  
 40 Trp Leu Ser  
 625

## (2) INFORMATION FOR SEQ ID NO:29:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1689 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: DNA (genomic)

55 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..1689  
 (D) OTHER INFORMATION: /product= "S8"

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GCA GGC GAT ACC GCA AAT ATT GGC GAC GGT GAT TTC TTG CCA CCG TAC 48  
 Ala Gly Asp Thr Ala Asn Ile Gly Asp Gly Asp Phe Leu Pro Pro Tyr  
 1 5 10 15

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5  | AAC | GAT | CTA | CTA | CTC | GGT | TAC | TGG | GAT | AAA | CTT | GAG | TTA | CGC | CTA | TAC | 34  |
|    | Asn | Asp | Val | Leu | Leu | Gly | Tyr | Trp | Asp | Lys | Leu | Glu | Leu | Arg | Leu | Tyr |     |
|    |     |     | 20  |     |     |     |     |     | 25  |     |     |     |     | 30  |     |     |     |
| 10 | AAC | CTG | CGC | CAC | AAT | CTG | AGT | CTG | GAT | GGT | CAA | CCG | CTA | AAT | CTG | CCA | 144 |
|    | Asn | Leu | Arg | His | Asn | Leu | Ser | Leu | Asp | Gly | Gln | Pro | Leu | Asn | Leu | Pro |     |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |
| 15 | CTG | TAT | GCC | ACG | CCG | GTA | GAC | CCG | AAA | ACC | CTG | CAA | CGC | CAG | CAA | GCC | 132 |
|    | Leu | Tyr | Ala | Thr | Pro | Val | Asp | Pro | Lys | Thr | Leu | Gln | Arg | Gln | Gln | Ala |     |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |     |
| 20 | GGA | GGG | GAC | GGT | ACA | GGC | AGT | AGT | CCG | GCT | GGT | GGT | CAA | GGC | AGT | GTT | 240 |
|    | Gly | Gly | Asp | Gly | Thr | Gly | Ser | Ser | Pro | Ala | Gly | Gly | Gln | Gly | Ser | Val |     |
|    |     | 65  |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |     |
| 25 | CAG | GGC | TGG | CGC | TAT | CCG | TTA | TTG | GTA | GAA | CGC | GCC | CGC | TCT | GCC | GTG | 288 |
|    | Gln | Gly | Trp | Arg | Tyr | Pro | Leu | Leu | Val | Glu | Arg | Ala | Arg | Ser | Ala | Val |     |
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |     |
| 30 | AGT | TTG | TTG | ACT | CAG | TTC | GGC | AAC | AGC | TTA | CAA | ACA | ACG | TTA | GAA | CAT | 336 |
|    | Ser | Leu | Leu | Thr | Gln | Phe | Gly | Asn | Ser | Leu | Gln | Thr | Thr | Leu | Glu | His |     |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |     |
| 35 | CAG | GAT | AAT | GAA | AAA | ATG | ACG | ATA | CTG | TTG | CAG | ACT | CAA | CAG | GAA | GCC | 384 |
|    | Gln | Asp | Asn | Glu | Lys | Met | Thr | Ile | Leu | Leu | Gln | Thr | Gln | Gln | Glu | Ala |     |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
| 40 | ATC | CTG | AAA | CAT | CAG | CAC | GAT | ATA | CAA | CAA | AAT | AAT | CTA | AAA | GGA | TTA | 432 |
|    | Ile | Leu | Lys | His | Gln | His | Asp | Ile | Gln | Gln | Asn | Asn | Leu | Lys | Gly | Leu |     |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |     |
| 45 | CAA | CAC | AGC | CTG | ACC | GCA | TTA | CAG | GCT | AGC | CGT | GAT | GGC | GAC | ACA | TTG | 480 |
|    | Gln | His | Ser | Leu | Thr | Ala | Leu | Gln | Ala | Ser | Arg | Asp | Gly | Asp | Thr | Leu |     |
|    |     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |     |
| 50 | CGG | CAA | AAA | CAT | TAC | AGC | GAC | CTG | ATT | AAC | GGT | GGT | CTA | TCT | GCG | GCA | 528 |
|    | Arg | Gln | Lys | His | Tyr | Ser | Asp | Leu | Ile | Asn | Gly | Gly | Leu | Ser | Ala | Ala |     |
|    |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |     |
| 55 | GAA | ATC | GCC | GGT | CTG | ACA | CTA | CGC | AGC | ACC | GCC | ATG | ATT | ACC | AAT | GGC | 576 |
|    | Glu | Ile | Ala | Gly | Leu | Thr | Leu | Arg | Ser | Thr | Ala | Met | Ile | Thr | Asn | Gly |     |
|    |     |     | 180 |     |     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |
| 60 | GTT | GCA | ACG | GGA | TTG | CTG | ATT | GCC | GGC | GGA | ATC | GCC | AAC | GCG | GTA | CCT | 624 |
|    | Val | Ala | Thr | Gly | Leu | Leu | Ile | Ala | Gly | Gly | Ile | Ala | Asn | Ala | Val | Pro |     |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |
| 65 | AAC | GTC | TTC | GGG | CTG | GCT | AAC | GGT | GGA | TCG | GAA | TGG | GGA | GCG | CCA | TTA | 672 |
|    | Asn | Val | Phe | Gly | Leu | Ala | Asn | Gly | Gly | Ser | Glu | Trp | Gly | Ala | Pro | Leu |     |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |     |
| 70 | ATT | GGC | TCC | GGG | CAA | GCA | ACC | CAA | GTT | GGC | GCC | GGC | ATC | CAG | GAT | CAG | 720 |
|    | Ile | Gly | Ser | Gly | Gln | Ala | Thr | Gln | Val | Gly | Ala | Gly | Ile | Gln | Asp | Gln |     |
|    |     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |     |
| 75 | AGC | GCG | GGC | ATT | TCA | GAA | GTG | ACA | GCA | GGC | TAT | CAG | CGT | CGT | CAG | GAA | 768 |
|    | Ser | Ala | Gly | Ile | Ser | Glu | Val | Thr | Ala | Gly | Tyr | Gln | Arg | Arg | Gln | Glu |     |
|    |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |     |
| 80 | GAA | TGG | GCA | TTG | CAA | CGG | GAT | ATT | GCT | GAT | AAC | GAA | ATA | ACC | CAA | CTG | 816 |
|    | Glu | Trp | Ala | Gln | Arg | Arg | Asp | Ile | Ala | Asp | Asn | Glu | Ile | Thr | Gln | Leu |     |
|    |     |     | 260 |     |     |     |     |     | 265 |     |     |     |     | 270 |     |     |     |
| 85 | GAT | GCC | CAG | ATA | CAA | AGC | CTG | CAA | GAG | CAA | ATC | ACG | ATG | GCA | CAA | AAA | 864 |
|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | Asp | Ala | Gln | Ile | Gln | Ser | Leu | Gln | Glu | Gln | Ile | Thr | Met | Ala | Gln | Lys |      |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |      |
| 5  | CAG | ATC | ACG | CTC | TCT | GAA | ACC | GAA | CAA | GCG | AAT | GCC | CAA | GCG | ATT | TAT | 912  |
|    | Gln | Ile | Thr | Leu | Ser | Glu | Thr | Glu | Gln | Ala | Asn | Ala | Gln | Ala | Ile | Tyr |      |
|    |     |     | 290 |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |      |
| 10 | GAG | CTG | CAA | ACC | ACT | CGT | TTT | ACC | GGG | CAG | GCA | CTG | TAT | AAC | TGG | ATG | 960  |
|    | Asp | Leu | Gln | Thr | Thr | Arg | Phe | Thr | Gly | Gln | Ala | Leu | Tyr | Asn | Trp | Met |      |
|    | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |      |
| 15 | GCC | GGT | CGT | CTC | TCC | GCG | CTC | TAT | TAC | CAA | ATG | TAT | GAT | TCC | ACT | CTG | 1008 |
|    | Ala | Gly | Arg | Leu | Ser | Ala | Leu | Tyr | Tyr | Gln | Met | Tyr | Asp | Ser | Thr | Leu |      |
|    |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |      |
| 20 | CCA | ATC | TGT | CTC | CAG | CCA | AAA | GCC | GCA | TTA | GTA | CAG | GAA | TTA | GGC | GAG | 1056 |
|    | Pro | Ile | Cys |     | Gln | Pro | Lys | Ala | Ala | Leu | Val | Gln | Glu | Leu | Gly | Glu |      |
|    |     |     |     | 340 |     |     |     | 345 |     |     |     |     |     | 350 |     |     |      |
| 25 | CAA | GGG | CTG | TTA | GCA | GGA | GAA | GGT | TTA | AGT | TCA | GAG | CTA | CAG | AAA | CTG | 1152 |
|    | Gln | Gly | Leu | Leu | Ala | Gly | Glu | Gly | Leu | Ser | Ser | Glu | Leu | Gln | Lys | Leu |      |
|    |     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |      |
| 30 | GAT | GCC | ATC | TGG | CTT | GCA | CGT | GGT | GGT | ATT | GGG | CTA | GAA | GCC | ATC | CGC | 1200 |
|    | Asp | Ala | Ile | Trp | Leu | Ala | Arg | Gly | Gly | Ile | Gly | Leu | Glu | Ala | Ile | Arg |      |
|    | 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |      |
| 35 | ACC | GTG | TCG | CTG | GAT | ACC | CTG | TTT | GGC | ACA | GGG | ACG | TTA | AGT | GAA | AAT | 1248 |
|    | Thr | Val | Ser | Leu | Asp | Thr | Leu | Phe | Gly | Thr | Gly | Thr | Leu | Ser | Glu | Asn |      |
|    |     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |      |
| 40 | ATC | AAT | AAA | GTG | CTT | AAC | GGG | GAA | ACG | GTA | TCT | CCA | TCC | GGT | GGC | GTC | 1296 |
|    | Ile | Asn | Lys | Val | Leu | Asn | Gly | Glu | Thr | Val | Ser | Pro | Ser | Gly | Gly | Val |      |
|    |     |     |     | 420 |     |     |     | 425 |     |     |     |     |     | 430 |     |     |      |
| 45 | ACT | CTG | GCG | CTG | ACA | GGG | GAT | ATC | TTC | CAA | GCA | ACA | CTG | GAT | TTG | AGT | 1344 |
|    | Thr | Leu | Ala | Leu | Thr | Gly | Asp | Ile | Phe | Gln | Ala | Thr | Leu | Asp | Leu | Ser |      |
|    |     |     | 435 |     |     |     | 440 |     |     |     |     | 445 |     |     |     |     |      |
| 50 | CAG | CTA | GGT | TTG | GAT | AAC | TCT | TAC | AAC | TTG | GGT | AAC | GAG | AAG | AAA | CGT | 1392 |
|    | Gln | Leu | Gly | Leu | Asp | Asn | Ser | Tyr | Asn | Leu | Gly | Asn | Glu | Lys | Lys | Arg |      |
|    |     | 450 |     |     |     | 455 |     |     |     |     |     | 460 |     |     |     |     |      |
| 55 | CGT | ATT | AAA | CGT | ATC | GCC | GTC | ACC | CTG | CCA | ACA | CTT | CTG | GGG | CCA | TAT | 1440 |
|    | Arg | Ile | Lys | Arg | Ile | Ala | Val | Thr | Leu | Pro | Thr | Leu | Leu | Gly | Pro | Tyr |      |
|    | 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |      |
| 60 | CAA | GAT | CTT | GAA | GCC | ACA | CTG | GTA | ATG | GGT | GCG | GAA | ATC | GCC | GCC | TTA | 1488 |
|    | Gln | Asp | Leu | Glu | Ala | Thr | Leu | Val | Met | Gly | Ala | Glu | Ile | Ala | Ala | Leu |      |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |      |
| 65 | TCA | CAC | GGT | GTG | AAT | GAC | GGA | GGC | CGG | TTT | GTT | ACC | GAC | TTT | AAC | GAC | 1536 |
|    | Ser | His | Gly | Val | Asn | Asp | Gly | Gly | Arg | Phe | Val | Thr | Asp | Phe | Asn | Asp |      |
|    |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |      |
| 70 | AGC | CGT | TTT | CTG | CCT | TTT | GAA | GGT | CGA | GAT | GCA | ACA | ACC | GGC | ACA | CTG | 1584 |
|    | Ser | Arg | Phe | Leu | Pro | Phe | Glu | Gly | Arg | Asp | Ala | Thr | Thr | Gly | Thr | Leu |      |
|    |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |      |
| 75 | GAG | CTC | AAT | ATT | TTC | CAT | GCG | GGT | AAA | GAG | GGA | ACG | CAA | CAC | GAG | TTG | 1632 |
|    | Glu | Leu | Asn | Ile | Phe | His | Ala | Gly | Lys | Glu | Gly | Thr | Gln | His | Glu | Leu |      |
|    |     | 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |      |

CTC GCG AAT CTG AGT GAC ATC ATT GTC CAT CTG AAT TAC ATC ATT CGA 1680  
 Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr Ile Ile Arg 550  
 545 550 555 560

5

GAC GCG TAA  
 Asp Ala \*

1689

10

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH: 563 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Ala Gly Asp Thr Ala Asn Ile Gly Asp Gly Asp Phe Leu Pro Pro Tyr  
 1 5 10 15  
 Asn Asp Val Leu Leu Gly Tyr Trp Asp Lys Leu Glu Leu Arg Leu Tyr  
 20 25 30  
 Asn Leu Arg His Asn Leu Ser Leu Asp Gly Gln Pro Leu Asn Leu Pro  
 35 40 45  
 Leu Tyr Ala Thr Pro Val Asp Pro Lys Thr Leu Gln Arg Gln Gln Ala  
 50 55 60  
 Gly Gly Asp Gly Thr Gly Ser Ser Pro Ala Gly Gly Gln Gly Ser Val  
 65 70 75 80  
 Gln Gly Trp Arg Tyr Pro Leu Leu Val Glu Arg Ala Arg Ser Ala Val  
 85 90 95  
 Ser Leu Leu Thr Gln Phe Gly Asn Ser Leu Gln Thr Thr Leu Glu His  
 100 105 110  
 Gln Asp Asn Glu Lys Met Thr Ile Leu Leu Gln Thr Gln Gln Glu Ala  
 115 120 125  
 Ile Leu Lys His Gln His Asp Ile Gln Gln Asn Asn Leu Lys Gly Leu  
 130 135 140  
 Gln His Ser Leu Thr Ala Leu Gln Ala Ser Arg Asp Gly Asp Thr Leu  
 145 150 155 160  
 Arg Gln Lys His Tyr Ser Asp Leu Ile Asn Gly Gly Leu Ser Ala Ala  
 165 170 175  
 Glu Ile Ala Gly Leu Thr Leu Arg Ser Thr Ala Met Ile Thr Asn Gly  
 180 185 190  
 Val Ala Thr Gly Leu Leu Ile Ala Gly Gly Ile Ala Asn Ala Val Pro  
 195 200 205  
 Asn Val Phe Gly Leu Ala Asn Gly Gly Ser Glu Trp Gly Ala Pro Leu  
 210 215 220  
 Ile Gly Ser Gly Gln Ala Thr Gln Val Gly Ala Gly Ile Gln Asp Gln

225                      230                      235                      240  
 Ser Ala Gly Ile Ser Glu Val Thr Ala Gly Tyr Gln Arg Arg Gln Glu  
                                  245                      250                      255  
 5    Glu Trp Ala Leu Gln Arg Asp Ile Ala Asp Asn Glu Ile Thr Gln Leu  
                                  260                      265                      270  
 10   Asp Ala Gln Ile Gln Ser Leu Gln Glu Gln Ile Thr Met Ala Gln Lys  
                                  275                      280                      285  
 Gln Ile Thr Leu Ser Glu Thr Glu Gln Ala Asn Ala Gln Ala Ile Tyr  
                                  290                      295                      300  
 15   Asp Leu Gln Thr Thr Arg Phe Thr Gly Gln Ala Leu Tyr Asn Trp Met  
                                  305                      310                      315                      320  
 Ala Gly Arg Leu Ser Ala Leu Tyr Tyr Gln Met Tyr Asp Ser Thr Leu  
                                  325                      330                      335  
 20   Pro Ile Cys Leu Gln Pro Lys Ala Ala Leu Val Gln Glu Leu Gly Glu  
                                  340                      345                      350  
 25   Lys Glu Ser Asp Ser Leu Phe Gln Val Pro Val Trp Asn Asp Leu Trp  
                                  355                      360                      365  
 Gln Gly Leu Leu Ala Gly Glu Gly Leu Ser Ser Glu Leu Gln Lys Leu  
                                  370                      375                      380  
 30   Asp Ala Ile Trp Leu Ala Arg Gly Gly Ile Gly Leu Glu Ala Ile Arg  
                                  385                      390                      395                      400  
 Thr Val Ser Leu Asp Thr Leu Phe Gly Thr Gly Thr Leu Ser Glu Asn  
                                  405                      410                      415  
 35   Ile Asn Lys Val Leu Asn Gly Glu Thr Val Ser Pro Ser Gly Gly Val  
                                  420                      425                      430  
 40   Thr Leu Ala Leu Thr Gly Asp Ile Phe Gln Ala Thr Leu Asp Leu Ser  
                                  435                      440                      445  
 Gln Leu Gly Leu Asp Asn Ser Tyr Asn Leu Gly Asn Glu Lys Lys Arg  
                                  450                      455                      460  
 45   Arg Ile Lys Arg Ile Ala Val Thr Leu Pro Thr Leu Leu Gly Pro Tyr  
                                  465                      470                      475                      480  
 Gln Asp Leu Glu Ala Thr Leu Val Met Gly Ala Glu Ile Ala Ala Leu  
                                  485                      490                      495  
 50   Ser His Gly Val Asn Asp Gly Gly Arg Phe Val Thr Asp Phe Asn Asp  
                                  500                      505                      510  
 55   Ser Arg Phe Leu Pro Phe Glu Gly Arg Asp Ala Thr Thr Gly Thr Leu  
                                  515                      520                      525  
 Glu Leu Asn Ile Phe His Ala Gly Lys Glu Gly Thr Gln His Glu Leu  
                                  530                      535                      540  
 60   Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr Ile Ile Arg  
                                  545                      550                      555                      560  
 Asp Ala    \*

65

## (2) INFORMATION FOR SEQ ID NO:31:

## (i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 4458 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

10

## (ix) FEATURE:

15

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..4458

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

|    |                                                                 |     |
|----|-----------------------------------------------------------------|-----|
| 20 | ATG CAG GAT TCA CCA GAA GTA TCG ATT ACA ACG CTG TCA CTT CCC AAA | 48  |
|    | Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Leu Ser Leu Pro Lys |     |
|    | 1 5 10 15                                                       |     |
| 25 | GGT GGC GGT GCT ATC AAT GGC ATG GGA GAA GCA CTG AAT GCT GCC GGC | 96  |
|    | Gly Gly Gly Ala Ile Asn Gly Met Gly Glu Ala Leu Asn Ala Ala Gly |     |
|    | 20 25 30                                                        |     |
| 30 | CCT GAT GGA ATG GCC TCC CTA TCT CTG CCA TTA CCC CTT TCG ACC GGC | 144 |
|    | Pro Asp Gly Met Ala Ser Leu Ser Leu Pro Leu Pro Leu Ser Thr Gly |     |
|    | 35 40 45                                                        |     |
| 35 | AGA GGG ACG GCT CCT GGA TTA TCG CTG ATT TAC AGC AAC AGT GCA GGT | 192 |
|    | Arg Gly Thr Ala Pro Gly Leu Ser Leu Ile Tyr Ser Asn Ser Ala Gly |     |
|    | 50 55 60                                                        |     |
| 40 | AAT GGG CCT TTC GGC ATC GGC TGG CAA TGC GGT GTT ATG TCC ATT AGC | 240 |
|    | Asn Gly Pro Phe Gly Ile Gly Trp Gln Cys Gly Val Met Ser Ile Ser |     |
|    | 65 70 75 80                                                     |     |
| 45 | CGA CGC ACC CAA CAT GGC ATT CCA CAA TAC GGT AAT GAC GAC ACG TTC | 288 |
|    | Arg Arg Thr Gln His Gly Ile Pro Gln Tyr Gly Asn Asp Asp Thr Phe |     |
|    | 85 90 95                                                        |     |
| 50 | CTA TCC CCA CAA GGC GAG GTC ATG AAT ATC GCC CTG AAT GAC CAA GGC | 336 |
|    | Leu Ser Pro Gln Gly Glu Val Met Asn Ile Ala Leu Asn Asp Gln Gly |     |
|    | 100 105 110                                                     |     |
| 55 | CAA CCT GAT ATC CGT CAA GAC GTT AAA ACG CTG CAA GGC GTT ACC TTG | 384 |
|    | Gln Pro Asp Ile Arg Gln Asp Val Lys Thr Leu Gln Gly Val Thr Leu |     |
|    | 115 120 125                                                     |     |
| 60 | CCA ATT TCC TAT ACC GTG ACC CGC TAT CAA GCC CGC CAG ATC CTG GAT | 432 |
|    | Pro Ile Ser Tyr Thr Val Thr Arg Tyr Gln Ala Arg Gln Ile Leu Asp |     |
|    | 130 135 140                                                     |     |
| 65 | TTC AGT AAA ATC GAA TAC TGG CAA CCT GCC TCC GGT CAA GAA GGA CGC | 480 |
|    | Phe Ser Lys Ile Glu Tyr Trp Gln Pro Ala Ser Gly Gln Glu Gly Arg |     |
|    | 145 150 155 160                                                 |     |
| 70 | GCT TTC TGG CTG ATA TCG ACA CCG GAC GGG CAT CTA CAC ATC TTA GGG | 528 |
|    | Ala Phe Trp Leu Ile Ser Thr Pro Asp Gly His Leu His Ile Leu Gly |     |
|    | 165 170 175                                                     |     |
|    | AAA ACC GCG CAG GCT TGT CTG GCA AAT CCG CAA AAT GAC CAA CAA ATC | 576 |
|    | Lys Thr Ala Gln Ala Cys Leu Ala Asn Pro Gln Asn Asp Gln Gln Ile |     |



|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | CCA | CTG | CGT | ACC | CTA | GGC | AAT | TTG | CAG | GAT | AAT | GCC | TCA | TTG | ATG | GAT | 1330 |
|    | Pro | Leu | Pro | Thr | Leu | Pro | Asn | Leu | Gln | Asp | Asn | Ala | Ser | Leu | Met | Asp |      |
|    | 450 |     |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |      |
| 5  | ATC | AAC | GGA | GAC | GGC | CAA | CTG | GAT | TGG | GTT | GTT | ACC | GCC | TCC | GGT | ATT | 1440 |
|    | Ile | Asn | Gly | Asp | Gly | Gln | Leu | Asp | Trp | Val | Val | Thr | Ala | Ser | Gly | Ile |      |
|    | 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |      |
| 10 | CGC | GGA | TAC | CAT | AGT | CAG | CAA | CCC | GAT | GGA | AAG | TGG | ACG | CAC | TTT | ACG | 1438 |
|    | Arg | Gly | Tyr | His | Ser | Gln | Gln | Pro | Asp | Gly | Lys | Trp | Thr | His | Phe | Thr |      |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |      |
| 15 | CCA | ATC | AAT | GCC | TTG | CCC | GTG | GAA | TAT | TTT | CAT | CCA | AGC | ATC | CAG | TTG | 1536 |
|    | Pro | Ile | Asn | Ala | Leu | Pro | Val | Glu | Tyr | Phe | His | Pro | Ser | Ile | Gln | Phe |      |
|    |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |      |
| 20 | GCT | GAC | CTT | ACC | GGG | GCA | GGC | TTA | TCT | GAT | TTA | GTG | TTG | ATC | GGG | CCG | 1584 |
|    | Ala | Asp | Leu | Thr | Gly | Ala | Gly | Leu | Ser | Asp | Leu | Val | Leu | Ile | Gly | Pro |      |
|    |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |      |
|    | AAA | AGC | GTG | CGT | CTA | TAT | GCC | AAC | CAG | CGA | AAC | GGC | TGG | CGT | AAA | GGA | 1632 |
|    | Lys | Ser | Val | Arg | Leu | Tyr | Ala | Asn | Gln | Arg | Asn | Gly | Trp | Arg | Lys | Gly |      |
|    | 530 |     |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |      |
| 25 | GAA | GAT | GTC | CCC | CAA | TCC | ACA | GGT | ATC | ACC | CTG | CCT | GTC | ACA | GGG | ACC | 1680 |
|    | Glu | Asp | Val | Pro | Gln | Ser | Thr | Gly | Ile | Thr | Leu | Pro | Val | Thr | Gly | Thr |      |
|    | 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |      |
| 30 | GAT | GCC | CGC | AAA | CTG | GTG | GCT | TTC | AGT | GAT | ATG | CTC | GGT | TCC | GGT | CAA | 1728 |
|    | Asp | Ala | Arg | Lys | Leu | Val | Ala | Phe | Ser | Asp | Met | Leu | Gly | Ser | Gly | Gln |      |
|    |     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |      |
| 35 | CAA | CAT | CTG | GTG | GAA | ATC | AAG | GGT | AAT | CGC | GTC | ACC | TGT | TGG | CCG | AAT | 1776 |
|    | Gln | His | Leu | Val | Glu | Ile | Lys | Gly | Asn | Arg | Val | Thr | Cys | Trp | Pro | Asn |      |
|    |     |     | 580 |     |     |     |     | 585 |     |     |     |     |     | 590 |     |     |      |
| 40 | CTA | GGG | CAT | GGC | CGT | TTC | GGT | CAA | CCA | CTA | ACT | CTG | TCA | GGA | TTT | AGC | 1824 |
|    | Leu | Gly | His | Gly | Arg | Phe | Gly | Gln | Pro | Leu | Thr | Leu | Ser | Gly | Phe | Ser |      |
|    |     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |      |
|    | CAG | CCC | GAA | AAT | AGC | TTC | AAT | CCC | GAA | CGG | CTG | TTT | CTG | GCG | GAT | ATC | 1872 |
|    | Gln | Pro | Glu | Asn | Ser | Phe | Asn | Pro | Glu | Arg | Leu | Phe | Leu | Ala | Asp | Ile |      |
|    | 610 |     |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |      |
| 45 | GAC | GGC | TCC | GGC | ACC | ACC | GAC | CTT | ATC | TAT | GCG | CAA | TCC | GGC | TCT | TTG | 1920 |
|    | Asp | Gly | Ser | Gly | Thr | Thr | Asp | Leu | Ile | Tyr | Ala | Gln | Ser | Gly | Ser | Leu |      |
|    | 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |      |
| 50 | CTC | ATT | TAT | CTC | AAC | CAA | AGT | GGT | AAT | CAG | TTT | GAT | GCC | CCG | TTG | ACA | 1968 |
|    | Leu | Ile | Tyr | Leu | Asn | Gln | Ser | Gly | Asn | Gln | Phe | Asp | Ala | Pro | Leu | Thr |      |
|    |     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |      |
| 55 | TTA | GCG | TTG | CCA | GAA | GGC | GTA | CAA | TTT | GAC | AAC | ACT | TGC | CAA | CTT | CAA | 2016 |
|    | Leu | Ala | Leu | Pro | Glu | Gly | Val | Gln | Phe | Asp | Asn | Thr | Cys | Gln | Leu | Gln |      |
|    |     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |      |
| 60 | GTC | GCC | GAT | ATT | CAG | GGA | TTA | GGG | ATA | GCC | AGC | TTG | ATT | CTG | ACT | GTG | 2064 |
|    | Val | Ala | Asp | Ile | Gln | Gly | Leu | Gly | Ile | Ala | Ser | Leu | Ile | Leu | Thr | Val |      |
|    |     |     | 675 |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |      |
|    | CCA | CAT | ATC | GCG | CCA | CAT | CAC | TGG | CGT | TGT | GAC | CTG | TCA | CTG | ACC | AAA | 2112 |
|    | Pro | His | Ile | Ala | Pro | His | His | Trp | Arg | Cys | Asp | Leu | Ser | Leu | Thr | Lys |      |
|    | 690 |     |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |      |
| 65 | CCC | TGG | TTG | TTG | AAT | GTA | ATG | AAC | AAT | AAC | CGG | GGC | GCA | CAT | CAC | ACG | 2160 |
|    | Pro | Trp | Leu | Leu | Asn | Val | Met | Asn | Asn | Asn | Arg | Gly | Ala | His | His | Thr |      |



|    | 705                                                                                                                                                    | 710 | 715 | 720 |  |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-----|-----|--|
| 5  | CTA CAT TAT CGT AGT TCC GCG CAA TTC TGG TTG GAT GAA AAA TTA CAG 2208<br>Leu His Tyr Arg Ser Ser Ala Gln Phe Trp Leu Asp Glu Lys Leu Gln<br>725 730 735 |     |     |     |  |
| 10 | CTC ACC AAA GCA GGC AAA TCT CCG GCT TGT TAT CTG CCG TTT CCA ATG 2256<br>Leu Thr Lys Ala Gly Lys Ser Pro Ala Cys Tyr Leu Pro Phe Pro Met<br>740 745 750 |     |     |     |  |
| 15 | CAT TTG CTA TGG TAT ACC GAA ATT CAG GAT GAA ATC AGC GGC AAC CCG 2304<br>His Leu Leu Trp Tyr Thr Glu Ile Gln Asp Glu Ile Ser Gly Asn Arg<br>755 760 765 |     |     |     |  |
| 20 | CTC ACC AGT GAA GTC AAC TAC AGC CAC GGC GTC TGG GAT GGT AAA GAG 2352<br>Leu Thr Ser Glu Val Asn Tyr Ser His Gly Val Trp Asp Gly Lys Glu<br>770 775 780 |     |     |     |  |
| 25 | CGG GAA TTC AGA GGA TTT GGC TGC ATC AAA CAG ACA GAT ACC ACA ACG 2400<br>Arg Glu Phe Arg Gly Phe Gly Cys Ile Lys Gln Thr Asp Thr Thr Thr<br>785 790 795 |     |     |     |  |
| 30 | TTT TCT CAC GGC ACC GCC CCC GAA CAG GCG GCA CCG TCG CTG AGT ATT 2448<br>Phe Ser His Gly Thr Ala Pro Glu Gln Ala Ala Pro Ser Leu Ser Ile<br>805 810 815 |     |     |     |  |
| 35 | AGC TGG TTT GCC ACC GGC ATG GAT GAA GTA GAC AGC CAA TTA GCT ACG 2496<br>Ser Trp Phe Ala Thr Gly Met Asp Glu Val Asp Ser Gln Leu Ala Thr<br>820 825 830 |     |     |     |  |
| 40 | GAA TAT TGG CAG GCA GAC ACG CAA GCT TAT AGC GGA TTT GAA ACC CGT 2544<br>Glu Tyr Trp Gln Ala Asp Thr Gln Ala Tyr Ser Gly Phe Glu Thr Arg<br>835 840 845 |     |     |     |  |
| 45 | TAT ACC GTC TGG GAT CAC ACC AAC CAG ACA GAC CAA GCA TTT ACC CCC 2592<br>Tyr Thr Val Trp Asp His Thr Asn Gln Thr Asp Gln Ala Phe Thr Pro<br>850 855 860 |     |     |     |  |
| 50 | AAT GAG ACA CAA CGT AAC TGG CTG ACG CGA GCG CTT AAA GGC CAA CTG 2640<br>Asn Glu Thr Gln Arg Asn Trp Leu Thr Arg Ala Leu Lys Gly Gln Leu<br>865 870 875 |     |     |     |  |
| 55 | CTA CGC ACT GAG CTC TAC GGT CTG GAC GGA ACA GAT AAG CAA ACA GTG 2688<br>Leu Arg Thr Glu Leu Tyr Gly Leu Asp Gly Thr Asp Lys Gln Thr Val<br>885 890 895 |     |     |     |  |
| 60 | CCT TAT ACC GTC AGT GAA TCG CGC TAT CAG GTA CGC TCT ATT CCC GTA 2736<br>Pro Tyr Thr Val Ser Glu Ser Arg Tyr Gln Val Arg Ser Ile Pro Val<br>900 905 910 |     |     |     |  |
| 65 | AAT AAA GAA ACT GAA TTA TCT GCC TGG GTG ACT GCT ATT GAA AAT CGC 2784<br>Asn Lys Glu Thr Glu Leu Ser Ala Trp Val Thr Ala Ile Glu Asn Arg<br>915 920 925 |     |     |     |  |
| 70 | AGC TAC CAC TAT GAA CGT ATC ATC ACT GAC CCA CAG TTC AGC CAG AGT 2832<br>Ser Tyr His Tyr Glu Arg Ile Ile Thr Asp Pro Gln Phe Ser Gln Ser<br>930 935 940 |     |     |     |  |
| 75 | ATC AAG TTG CAA CAC GAT ATC TTT GGT CAA TCA CTG CAA AGT GTC GAT 2880<br>Ile Lys Leu Gln His Asp Ile Phe Gly Gln Ser Leu Gln Ser Val Asp<br>945 950 955 |     |     |     |  |
| 80 | ATT GCC TGG CCG CGC CGC GAA AAA CCA GCA GTG AAT CCC TAC CCG CCT 2928<br>Ile Ala Trp Pro Arg Arg Glu Lys Pro Ala Val Asn Pro Tyr Pro Pro<br>965 970 975 |     |     |     |  |

|    |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |      |      |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-----|------|------|
|    | ACT  | CTG  | CCG  | GAA  | ACG  | CTA  | TTT  | GAC  | AGC  | AGC  | TAT  | GAT  | GAT  | CAA  | CAA | TAA  | 3276 |
|    | Thr  | Leu  | Pro  | Glu  | Thr  | Leu  | Phe  | Asp  | Ser  | Ser  | Tyr  | Asp  | Asp  | Gln  | Gln | Gln  |      |
|    |      |      |      | 380  |      |      |      |      | 385  |      |      |      |      | 990  |     |      |      |
| 5  | CTA  | TTA  | CGT  | CTG  | GTG  | AGA  | CAA  | AAA  | AAT  | AGC  | TGG  | CAT  | CAC  | CTG  | ACT | GAT  | 3024 |
|    | Leu  | Leu  | Arg  | Leu  | Val  | Arg  | Gln  | Lys  | Asn  | Ser  | Trp  | His  | His  | Leu  | Thr | Asp  |      |
|    |      |      | 995  |      |      |      |      | 1000 |      |      |      |      | 1005 |      |     |      |      |
| 10 | CCG  | GAA  | AAC  | TGG  | CGA  | TTA  | GGT  | TTA  | CCG  | AAT  | GCA  | CAA  | CGC  | CGT  | GAT | CTT  | 3072 |
|    | Gly  | Glu  | Asn  | Trp  | Arg  | Leu  | Gly  | Leu  | Pro  | Asn  | Ala  | Gln  | Arg  | Arg  | Asp | Val  |      |
|    |      | 1010 |      |      |      |      | 1015 |      |      |      |      | 1020 |      |      |     |      |      |
| 15 | TAT  | ACT  | TAT  | GAC  | CCG  | AGC  | AAA  | ATT  | CCA  | ACC  | GAA  | GGG  | ATT  | TCC  | CTT | GAA  | 3120 |
|    | Tyr  | Thr  | Tyr  | Asp  | Arg  | Ser  | Lys  | Ile  | Pro  | Thr  | Glu  | Gly  | Ile  | Ser  | Leu | Glu  |      |
|    | 1025 |      |      |      |      | 1030 |      |      |      |      | 1035 |      |      |      |     | 1040 |      |
| 20 | ATC  | TTG  | CTG  | AAA  | GAT  | GAT  | GGC  | CTG  | CTA  | GCA  | GAT  | GAA  | AAA  | GCG  | GCC | GTT  | 3168 |
|    | Ile  | Leu  | Leu  | Lys  | Asp  | Asp  | Gly  | Leu  | Leu  | Ala  | Asp  | Glu  | Lys  | Ala  | Ala | Val  |      |
|    |      |      |      | 1045 |      |      |      |      |      | 1050 |      |      |      | 1055 |     |      |      |
|    | TAT  | CTG  | GGA  | CAA  | CAA  | CAG  | ACG  | TTT  | TAC  | ACC  | GCC  | GGT  | CAA  | GCG  | GAA | GTC  | 3216 |
|    | Tyr  | Leu  | Gly  | Gln  | Gln  | Gln  | Thr  | Phe  | Tyr  | Thr  | Ala  | Gly  | Gln  | Ala  | Glu | Val  |      |
|    |      |      | 1060 |      |      |      |      | 1065 |      |      |      |      | 1070 |      |     |      |      |
| 25 | ACT  | CTA  | GAA  | AAA  | CCC  | ACG  | TTA  | CAA  | GCA  | CTG  | GTC  | GCG  | TTC  | CAA  | GAA | ACC  | 3264 |
|    | Thr  | Leu  | Glu  | Lys  | Pro  | Thr  | Leu  | Gln  | Ala  | Leu  | Val  | Ala  | Phe  | Gln  | Glu | Thr  |      |
|    |      |      | 1075 |      |      |      | 1080 |      |      |      |      |      | 1085 |      |     |      |      |
| 30 | GCC  | ATG  | ATG  | GAC  | GAT  | ACC  | TCA  | TTA  | CAG  | GCG  | TAT  | GAA  | GGC  | GTG  | ATT | GAA  | 3312 |
|    | Ala  | Met  | Met  | Asp  | Asp  | Thr  | Ser  | Leu  | Gln  | Ala  | Tyr  | Glu  | Gly  | Val  | Ile | Glu  |      |
|    |      | 1090 |      |      |      |      | 1095 |      |      |      |      | 1100 |      |      |     |      |      |
| 35 | GAG  | CAA  | GAG  | TTG  | AAT  | ACC  | GCG  | CTG  | ACA  | CAG  | GCC  | GGT  | TAT  | CAG  | CAA | GTC  | 3360 |
|    | Glu  | Gln  | Glu  | Leu  | Asn  | Thr  | Ala  | Leu  | Thr  | Gln  | Ala  | Gly  | Tyr  | Gln  | Gln | Val  |      |
|    | 1105 |      |      |      | 1110 |      |      |      |      | 1115 |      |      |      | 1120 |     |      |      |
| 40 | GCG  | CGG  | TTG  | TTT  | AAT  | ACC  | AGA  | TCA  | GAA  | AGC  | CCG  | GTA  | TGG  | GCG  | GCA | CGG  | 3408 |
|    | Ala  | Arg  | Leu  | Phe  | Asn  | Thr  | Arg  | Ser  | Glu  | Ser  | Pro  | Val  | Trp  | Ala  | Ala | Arg  |      |
|    |      |      |      | 1125 |      |      |      | 1130 |      |      |      |      | 1135 |      |     |      |      |
|    | CAA  | GGT  | TAT  | ACC  | GAT  | TAC  | GGT  | GAC  | GCC  | GCA  | CAG  | TTC  | TGG  | CGG  | CCT | CAG  | 3456 |
|    | Gln  | Gly  | Tyr  | Thr  | Asp  | Tyr  | Gly  | Asp  | Ala  | Ala  | Gln  | Phe  | Trp  | Arg  | Pro | Gln  |      |
|    |      |      | 1140 |      |      |      | 1145 |      |      |      |      | 1150 |      |      |     |      |      |
| 45 | GCT  | CAG  | CGT  | AAC  | TCG  | TTG  | CTG  | ACA  | GGG  | AAA  | ACC  | ACA  | CTG  | ACC  | TGG | GAT  | 3504 |
|    | Ala  | Gln  | Arg  | Asn  | Ser  | Leu  | Leu  | Thr  | Gly  | Lys  | Thr  | Thr  | Leu  | Thr  | Trp | Asp  |      |
|    |      |      | 1155 |      |      |      | 1160 |      |      |      |      | 1165 |      |      |     |      |      |
| 50 | ACC  | CAT  | CAT  | TGT  | GTA  | ATA  | ATA  | CAG  | ACT  | CAA  | GAT  | GCC  | GCT  | GGA  | TTA | ACG  | 3552 |
|    | Thr  | His  | His  | Cys  | Val  | Ile  | Ile  | Gln  | Thr  | Gln  | Asp  | Ala  | Ala  | Gly  | Leu | Thr  |      |
|    |      | 1170 |      |      |      | 1175 |      |      |      |      | 1180 |      |      |      |     |      |      |
| 55 | ACG  | CAA  | GCC  | CAT  | TAC  | GAT  | TAT  | CGT  | TTC  | CTT  | ACA  | CCG  | GTA  | CAA  | CTG | ACA  | 3600 |
|    | Thr  | Gln  | Ala  | His  | Tyr  | Asp  | Tyr  | Arg  | Phe  | Leu  | Thr  | Pro  | Val  | Gln  | Leu | Thr  |      |
|    | 1185 |      |      |      | 1190 |      |      |      | 1195 |      |      |      | 1200 |      |     |      |      |
| 60 | GAT  | ATT  | AAT  | GAT  | AAT  | CAA  | CAT  | ATT  | GTG  | ACT  | CTG  | GAC  | GCG  | CTA  | GGT | CGC  | 3648 |
|    | Asp  | Ile  | Asn  | Asp  | Asn  | Gln  | His  | Ile  | Val  | Thr  | Leu  | Asp  | Ala  | Leu  | Gly | Arg  |      |
|    |      |      |      | 1205 |      |      |      | 1210 |      |      |      | 1215 |      |      |     |      |      |
|    | GTA  | ACC  | ACC  | AGC  | CGG  | TTC  | TGG  | GGC  | ACA  | GAG  | GCA  | GGA  | CAA  | GCC  | GCA | GGC  | 3696 |
|    | Val  | Thr  | Thr  | Ser  | Arg  | Phe  | Trp  | Gly  | Thr  | Glu  | Ala  | Gly  | Gln  | Ala  | Ala | Gly  |      |
|    |      |      |      | 1220 |      |      |      | 1225 |      |      |      | 1230 |      |      |     |      |      |
| 65 | TAT  | TCC  | AAC  | CAG  | CCC  | TTC  | ACA  | CCA  | CCG  | GAC  | TCC  | GTA  | GAT  | AAA  | GCG | CTG  | 3744 |
|    | Tyr  | Ser  | Asn  | Gln  | Pro  | Phe  | Thr  | Pro  | Pro  | Asp  | Ser  | Val  | Asp  | Lys  | Ala | Leu  |      |

|    | 1235                                                                                                                                                           | 1240 | 1245 |  |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------|--|
| 5  | GCA TTA ACC GGC GCA CTC CCT GTT GCC CAA TGT TTA GTC TAT GCC GTT 3732<br>Ala Leu Thr Gly Ala Leu Pro Val Ala Gln Cys Leu Val Tyr Ala Val<br>1250 1255 1260      |      |      |  |
| 10 | GAT AGC TGG ATG CCG TCG TTA TCT TTG TCT CAG CTT TCT CAG TCA CAA 3840<br>Asp Ser Trp Met Pro Ser Leu Ser Leu Ser Gln Leu Ser Gln Ser Gln<br>1265 1270 1275 1280 |      |      |  |
| 15 | GAA GAG GCA GAA GCG CTA TGG GCG CAA CTG CGT GCC GCT CAT ATG ATT 3888<br>Glu Glu Ala Glu Ala Leu Trp Ala Gln Leu Arg Ala Ala His Met Ile<br>1285 1290 1295      |      |      |  |
| 20 | ACC GAA GAT GGG AAA GTG TGT GCG TTA AGC GGG AAA CGA GGA ACA AGC 3936<br>Thr Glu Asp Gly Lys Val Cys Ala Leu Ser Gly Lys Arg Gly Thr Ser<br>1300 1305 1310      |      |      |  |
| 25 | CAT CAG AAC CTG ACG ATT CAA CTT ATT TCG CTA TTG GCA AGT ATT CCC 3984<br>His Gln Asn Leu Thr Ile Gln Leu Ile Ser Leu Leu Ala Ser Ile Pro<br>1315 1320 1325      |      |      |  |
| 30 | CGT TTA CCG CCA CAT GTA CTG GGG ATC ACC ACT GAT CGC TAT GAT AGC 4032<br>Arg Leu Pro Pro His Val Leu Gly Ile Thr Thr Asp Arg Tyr Asp Ser<br>1330 1335 1340      |      |      |  |
| 35 | GAT CCG CAA CAG CAG CAC CAA CAG ACG GTG AGC TTT AGT GAC GGT TTT 4080<br>Asp Pro Gln Gln Gln His Gln Gln Thr Val Ser Phe Ser Asp Gly Phe<br>1345 1350 1355 1360 |      |      |  |
| 40 | GGC CGG TTA CTC CAG AGT TCA GCT CGT CAT GAG TCA GGT GAT GCC TGG 4128<br>Gly Arg Leu Leu Gln Ser Ser Ala Arg His Glu Ser Gly Asp Ala Trp<br>1365 1370 1375      |      |      |  |
| 45 | CAA CGT AAA GAG GAT GGC GGG CTG GTC GTG GAT GCA AAT GGC GTT CTG 4176<br>Gln Arg Lys Glu Asp Gly Gly Leu Val Val Asp Ala Asn Gly Val Leu<br>1380 1385 1390      |      |      |  |
| 50 | GTC AGT GCC CCT ACA GAC ACC CGA TGG GCC GTT TCC GGT CGC ACA GAA 4224<br>Val Ser Ala Pro Thr Asp Thr Arg Trp Ala Val Ser Gly Arg Thr Glu<br>1395 1400 1405      |      |      |  |
| 55 | TAT GAC GAC AAA GGC CAA CCT GTG CGT ACT TAT CAA CCC TAT TTT CTA 4272<br>Tyr Asp Asp Lys Gly Gln Pro Val Arg Thr Tyr Gln Pro Tyr Phe Leu<br>1410 1415 1420      |      |      |  |
| 60 | AAT GAC TGG CGT TAC GTT AGT GAT GAC AGC GCA CGA GAT GAC CTG TTT 4320<br>Asn Asp Trp Arg Tyr Val Ser Asp Asp Ser Ala Arg Asp Asp Leu Phe<br>1425 1430 1435 1440 |      |      |  |
| 65 | GCC GAT ACC CAC CTT TAT GAT CCA TTG GGA CGG GAA TAC AAA GTC ATC 4368<br>Ala Asp Thr His Leu Tyr Asp Pro Leu Gly Arg Glu Tyr Lys Val Ile<br>1445 1450 1455      |      |      |  |
| 70 | ACT GCT AAG AAA TAT TTG CGA GAA AAG CTG TAC ACC CCG TGG TTT ATT 4416<br>Thr Ala Lys Lys Tyr Leu Arg Glu Lys Leu Tyr Thr Pro Trp Phe Ile<br>1460 1465 1470      |      |      |  |
| 75 | GTC AGT GAG GAT GAA AAC GAT ACA GCA TCA AGA ACC CCA TAG 4458<br>Val Ser Glu Asp Glu Asn Asp Thr Ala Ser Arg Thr Pro *<br>1475 1480 1485                        |      |      |  |

(2) INFORMATION FOR SEQ ID NO:32:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1486 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

10

Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Leu Ser Leu Pro Lys  
 1 5 10 15

15

Gly Gly Gly Ala Ile Asn Gly Met Gly Glu Ala Leu Asn Ala Ala Gly  
 20 25 30

Pro Asp Gly Met Ala Ser Leu Ser Leu Pro Leu Pro Leu Ser Thr Gly  
 35 40 45

20

Arg Gly Thr Ala Pro Gly Leu Ser Leu Ile Tyr Ser Asn Ser Ala Gly  
 50 55 60

25

Asn Gly Pro Phe Gly Ile Gly Trp Gln Cys Gly Val Met Ser Ile Ser  
 65 70 75 80

Arg Arg Thr Gln His Gly Ile Pro Gln Tyr Gly Asn Asp Asp Thr Phe  
 85 90 95

30

Leu Ser Pro Gln Gly Glu Val Met Asn Ile Ala Leu Asn Asp Gln Gly  
 100 105 110

Gln Pro Asp Ile Arg Gln Asp Val Lys Thr Leu Gln Gly Val Thr Leu  
 115 120 125

35

Pro Ile Ser Tyr Thr Val Thr Arg Tyr Gln Ala Arg Gln Ile Leu Asp  
 130 135 140

40

Phe Ser Lys Ile Glu Tyr Trp Gln Pro Ala Ser Gly Gln Glu Gly Arg  
 145 150 155 160

Ala Phe Trp Leu Ile Ser Thr Pro Asp Gly His Leu His Ile Leu Gly  
 165 170 175

45

Lys Thr Ala Gln Ala Cys Leu Ala Asn Pro Gln Asn Asp Gln Gln Ile  
 180 185 190

Ala Gln Trp Leu Leu Glu Glu Thr Val Thr Pro Ala Gly Glu His Val  
 195 200 205

50

Ser Tyr Gln Tyr Arg Ala Glu Asp Glu Ala His Cys Asp Asp Asn Glu  
 210 215 220

55

Lys Thr Ala His Pro Asn Val Thr Ala Gln Arg Tyr Leu Val Gln Val  
 225 230 235 240

Asn Tyr Gly Asn Ile Lys Pro Gln Ala Ser Leu Phe Val Leu Asp Asn  
 245 250 255

60

Ala Pro Pro Ala Pro Glu Glu Trp Leu Phe His Leu Val Phe Asp His  
 260 265 270

Gly Glu Arg Asp Thr Ser Leu His Thr Val Pro Thr Trp Asp Ala Gly  
 275 280 285

65

Thr Ala Gln Trp Ser Val Arg Pro Asp Ile Phe Ser Arg Tyr Glu Tyr

|    | 290     |     |         |         |         | 295     |         |         |         |         | 300     |         |         |         |         |         |
|----|---------|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 5  | Gly 305 | Phe | Glu     | Val     | Arg     | Thr 310 | Arg     | Arg     | Leu     | Cys     | Gln 315 | Gln     | Val     | Leu     | Met     | Phe 320 |
|    | His     | Arg | Thr     | Ala     | Leu 325 | Met     | Ala     | Gly     | Glu     | Ala 330 | Ser     | Thr     | Asn     | Asp     | Ala 335 | Pro     |
| 10 | Glu     | Leu | Val     | Gly 340 | Arg     | Leu     | Ile     | Leu     | Glu 345 | Tyr     | Asp     | Lys     | Asn     | Ala 350 | Ser     | Val     |
|    | Thr     | Thr | Leu 355 | Ile     | Thr     | Ile     | Arg     | Gln 360 | Leu     | Ser     | His     | Glu     | Ser 365 | Asp     | Gly     | Arg     |
| 15 | Pro     | Val | Thr     | Gln     | Pro     | Pro     | Leu 375 | Glu     | Leu     | Ala     | Trp     | Gln     | Arg     | Phe     | Asp     | Leu     |
|    | Glu 385 | Lys | Ile     | Pro     | Thr     | Trp 390 | Gln     | Arg     | Phe     | Asp     | Ala 395 | Leu     | Asp     | Asn     | Phe     | Asn 400 |
| 20 | Ser     | Gln | Gln     | Arg     | Tyr 405 | Gln     | Leu     | Val     | Asp     | Leu 410 | Arg     | Gly     | Glu     | Gly     | Leu     | Pro     |
|    | Gly     | Met | Leu     | Tyr 420 | Gln     | Asp     | Arg     | Gly     | Ala 425 | Trp     | Trp     | Tyr     | Lys     | Ala 430 | Pro     | Gln     |
| 25 | Arg     | Gln | Glu     | Asp 435 | Gly     | Asp     | Ser     | Asn 440 | Ala     | Val     | Thr     | Tyr     | Asp 445 | Lys     | Ile     | Ala     |
| 30 | Pro     | Leu | Pro     | Thr     | Leu     | Pro     | Asn 455 | Leu     | Gln     | Asp     | Asn     | Ala 460 | Ser     | Leu     | Met     | Asp     |
|    | Ile 465 | Asn | Gly     | Asp     | Gly     | Gln 470 | Leu     | Asp     | Trp     | Val     | Val 475 | Thr     | Ala     | Ser     | Gly     | Ile 480 |
| 35 | Arg     | Gly | Tyr     | His 485 | Ser     | Gln     | Gln     | Pro     | Asp     | Gly 490 | Lys     | Trp     | Thr     | His     | Phe 495 | Thr     |
| 40 | Pro     | Ile | Asn     | Ala 500 | Leu     | Pro     | Val     | Glu     | Tyr 505 | Phe     | His     | Pro     | Ser     | Ile 510 | Gln     | Phe     |
|    | Ala     | Asp | Leu 515 | Thr     | Gly     | Ala     | Gly     | Leu 520 | Ser     | Asp     | Leu     | Val     | Leu 525 | Ile     | Gly     | Pro     |
| 45 | Lys     | Ser | Val     | Arg     | Leu     | Tyr     | Ala 535 | Asn     | Gln     | Arg     | Asn     | Gly 540 | Trp     | Arg     | Lys     | Gly     |
|    | Glu 545 | Asp | Val     | Pro     | Gln     | Ser 550 | Thr     | Gly     | Ile     | Thr     | Leu 555 | Pro     | Val     | Thr     | Gly     | Thr 560 |
| 50 | Asp     | Ala | Arg     | Lys 565 | Leu     | Val     | Ala     | Phe     | Ser     | Asp 570 | Met     | Leu     | Gly     | Ser     | Gly 575 | Gln     |
|    | Gln     | His | Leu     | Val 580 | Glu     | Ile     | Lys     | Gly     | Asn 585 | Arg     | Val     | Thr     | Cys     | Trp     | Pro     | Asn     |
| 55 | Leu     | Gly | His 595 | Gly     | Arg     | Phe     | Gly     | Gln     | Pro     | Leu     | Thr     | Leu     | Ser 605 | Gly     | Phe     | Ser     |
| 60 | Gln     | Pro | Glu     | Asn 610 | Ser     | Phe     | Asn 615 | Pro     | Glu     | Arg     | Leu     | Phe     | Leu     | Ala     | Asp     | Ile     |
|    | Asp 625 | Gly | Ser     | Gly     | Thr     | Thr 630 | Asp     | Leu     | Ile     | Tyr     | Ala 635 | Gln     | Ser     | Gly     | Ser     | Leu 640 |
| 65 | Leu     | Ile | Tyr     | Leu     | Asn     | Gln     | Ser     | Gly     | Asn     | Gln     | Phe     | Asp     | Ala     | Pro     | Leu     | Thr     |

|    | 645                                                             | 650 | 655 |
|----|-----------------------------------------------------------------|-----|-----|
|    | Leu Ala Leu Pro Glu Gly Val Gln Phe Asp Asn Thr Cys Gln Leu Gln |     |     |
| 5  | 650                                                             | 665 | 670 |
|    | Val Ala Asp Ile Gln Gly Leu Gly Ile Ala Ser Leu Ile Leu Thr Val |     |     |
|    | 675                                                             | 680 | 685 |
| 10 | Pro His Ile Ala Pro His His Trp Arg Cys Asp Leu Ser Leu Thr Lys |     |     |
|    | 690                                                             | 695 | 700 |
|    | Pro Trp Leu Leu Asn Val Met Asn Asn Asn Arg Gly Ala His His Thr |     |     |
|    | 705                                                             | 710 | 715 |
| 15 | Leu His Tyr Arg Ser Ser Ala Gln Phe Trp Leu Asp Glu Lys Leu Gln |     |     |
|    | 725                                                             | 730 | 735 |
|    | Leu Thr Lys Ala Gly Lys Ser Pro Ala Cys Tyr Leu Pro Phe Pro Met |     |     |
| 20 | 740                                                             | 745 | 750 |
|    | His Leu Leu Trp Tyr Thr Glu Ile Gln Asp Glu Ile Ser Gly Asn Arg |     |     |
|    | 755                                                             | 760 | 765 |
| 25 | Leu Thr Ser Glu Val Asn Tyr Ser His Gly Val Trp Asp Gly Lys Glu |     |     |
|    | 770                                                             | 775 | 780 |
|    | Arg Glu Phe Arg Gly Phe Gly Cys Ile Lys Gln Thr Asp Thr Thr Thr |     |     |
|    | 785                                                             | 790 | 795 |
| 30 | Phe Ser His Gly Thr Ala Pro Glu Gln Ala Ala Pro Ser Leu Ser Ile |     |     |
|    | 805                                                             | 810 | 815 |
|    | Ser Trp Phe Ala Thr Gly Met Asp Glu Val Asp Ser Gln Leu Ala Thr |     |     |
| 35 | 820                                                             | 825 | 830 |
|    | Glu Tyr Trp Gln Ala Asp Thr Gln Ala Tyr Ser Gly Phe Glu Thr Arg |     |     |
|    | 835                                                             | 840 | 845 |
| 40 | Tyr Thr Val Trp Asp His Thr Asn Gln Thr Asp Gln Ala Phe Thr Pro |     |     |
|    | 850                                                             | 855 | 860 |
|    | Asn Glu Thr Gln Arg Asn Trp Leu Thr Arg Ala Leu Lys Gly Gln Leu |     |     |
|    | 865                                                             | 870 | 875 |
| 45 | Leu Arg Thr Glu Leu Tyr Gly Leu Asp Gly Thr Asp Lys Gln Thr Val |     |     |
|    | 885                                                             | 890 | 895 |
|    | Pro Tyr Thr Val Ser Glu Ser Arg Tyr Gln Val Arg Ser Ile Pro Val |     |     |
| 50 | 900                                                             | 905 | 910 |
|    | Asn Lys Glu Thr Glu Leu Ser Ala Trp Val Thr Ala Ile Glu Asn Arg |     |     |
|    | 915                                                             | 920 | 925 |
| 55 | Ser Tyr His Tyr Glu Arg Ile Ile Thr Asp Pro Gln Phe Ser Gln Ser |     |     |
|    | 930                                                             | 935 | 940 |
|    | Ile Lys Leu Gln His Asp Ile Phe Gly Gln Ser Leu Gln Ser Val Asp |     |     |
|    | 945                                                             | 950 | 955 |
| 60 | Ile Ala Trp Pro Arg Arg Glu Lys Pro Ala Val Asn Pro Tyr Pro Pro |     |     |
|    | 965                                                             | 970 | 975 |
|    | Thr Leu Pro Glu Thr Leu Phe Asp Ser Ser Tyr Asp Asp Gln Gln Gln |     |     |
| 65 | 980                                                             | 985 | 990 |
|    | Leu Leu Arg Leu Val Arg Gln Lys Asn Ser Trp His His Leu Thr Asp |     |     |

|    | 995                                                             | 1000                    | 1005                |
|----|-----------------------------------------------------------------|-------------------------|---------------------|
|    | Gly Glu Asn Trp Arg Leu                                         | Gly Leu Pro Asn Ala     | Gln Arg Arg Asp Val |
|    | 1010                                                            | 1015                    | 1020                |
| 5  | Tyr Thr Tyr Asp Arg Ser Lys Ile Pro Thr                         | Glu Gly Ile Ser Leu Glu |                     |
|    | 1025                                                            | 1030                    | 1035 1040           |
|    | Ile Leu Leu Lys Asp Asp Gly Leu Leu Ala Asp                     | Glu Lys Ala Ala Val     |                     |
| 10 |                                                                 | 1045                    | 1050 1055           |
|    | Tyr Leu Gly Gln Gln Gln Thr Phe Tyr Thr                         | Ala Gly Gln Ala Glu Val |                     |
|    |                                                                 | 1060                    | 1065 1070           |
| 15 | Thr Leu Glu Lys Pro Thr Leu Gln Ala Leu Val                     | Ala Phe Gln Glu Thr     |                     |
|    |                                                                 | 1075                    | 1080 1085           |
|    | Ala Met Met Asp Asp Thr Ser Leu Gln Ala Tyr                     | Glu Gly Val Ile Glu     |                     |
| 20 |                                                                 | 1090                    | 1095 1100           |
|    | Glu Gln Glu Leu Asn Thr Ala Leu Thr Gln Ala Gly Tyr             | Gln Gln Val             |                     |
|    |                                                                 | 1105                    | 1110 1115 1120      |
|    | Ala Arg Leu Phe Asn Thr Arg Ser Glu Ser Pro Val Trp             | Ala Ala Arg             |                     |
| 25 |                                                                 | 1125                    | 1130 1135           |
|    | Gln Gly Tyr Thr Asp Tyr Gly Asp Ala Ala Gln Phe Trp             | Arg Pro Gln             |                     |
|    |                                                                 | 1140                    | 1145 1150           |
| 30 | Ala Gln Arg Asn Ser Leu Leu Thr Gly Lys Thr Thr                 | Leu Thr Trp Asp         |                     |
|    |                                                                 | 1155                    | 1160 1165           |
|    | Thr His His Cys Val Ile Ile Gln Thr Gln Asp Ala Ala Gly         | Leu Thr                 |                     |
| 35 |                                                                 | 1170                    | 1175 1180           |
|    | Thr Gln Ala His Tyr Asp Tyr Arg Phe Leu Thr Pro Val Gln         | Leu Thr                 |                     |
|    |                                                                 | 1185                    | 1190 1195 1200      |
|    | Asp Ile Asn Asp Asn Gln His Ile Val Thr Leu Asp Ala Leu         | Gly Arg                 |                     |
| 40 |                                                                 | 1205                    | 1210 1215           |
|    | Val Thr Thr Ser Arg Phe Trp Gly Thr Glu Ala Gly Gln Ala Ala Gly |                         |                     |
|    |                                                                 | 1220                    | 1225 1230           |
| 45 | Tyr Ser Asn Gln Pro Phe Thr Pro Pro Asp Ser Val Asp Lys Ala Leu |                         |                     |
|    |                                                                 | 1235                    | 1240 1245           |
|    | Ala Leu Thr Gly Ala Leu Pro Val Ala Gln Cys Leu Val Tyr Ala Val |                         |                     |
| 50 |                                                                 | 1250                    | 1255 1260           |
|    | Asp Ser Trp Met Pro Ser Leu Ser Leu Ser Gln Leu Ser Gln Ser Gln |                         |                     |
|    |                                                                 | 1265                    | 1270 1275 1280      |
|    | Glu Glu Ala Glu Ala Leu Trp Ala Gln Leu Arg Ala Ala His Met Ile |                         |                     |
| 55 |                                                                 | 1285                    | 1290 1295           |
|    | Thr Glu Asp Gly Lys Val Cys Ala Leu Ser Gly Lys Arg Gly Thr Ser |                         |                     |
|    |                                                                 | 1300                    | 1305 1310           |
| 60 | His Gln Asn Leu Thr Ile Gln Leu Ile Ser Leu Leu Ala Ser Ile Pro |                         |                     |
|    |                                                                 | 1315                    | 1320 1325           |
|    | Arg Leu Pro Pro His Val Leu Gly Ile Thr Thr Asp Arg Tyr Asp Ser |                         |                     |
| 65 |                                                                 | 1330                    | 1335 1340           |
|    | Asp Pro Gln Gln Gln His Gln Gln Thr Val Ser Phe Ser Asp Gly Phe |                         |                     |

1345                      1350                      1355                      1360  
 Gly Arg Leu Leu Gln Ser Ser Ala Arg His Glu Ser Gly Asp Ala Trp  
                                  1365                      1370                      1375  
 5    Gln Arg Lys Glu Asp Gly Gly Leu Val Val Asp Ala Asn Gly Val Leu  
                                  1380                      1385                      1390  
 10   Val Ser Ala Pro Thr Asp Thr Arg Trp Ala Val Ser Gly Arg Thr Glu  
                                  1395                      1400                      1405  
     Tyr Asp Asp Lys Gly Gln Pro Val Arg Thr Tyr Gln Pro Tyr Phe Leu  
                                  1410                      1415                      1420  
 15   Asn Asp Trp Arg Tyr Val Ser Asp Asp Ser Ala Arg Asp Asp Leu Phe  
                                  1425                      1430                      1435                      1440  
     Ala Asp Thr His Leu Tyr Asp Pro Leu Gly Arg Glu Tyr Lys Val Ile  
                                  1445                      1450                      1455  
 20   Thr Ala Lys Lys Tyr Leu Arg Glu Lys Leu Tyr Thr Pro Trp Phe Ile  
                                  1460                      1465                      1470  
 25   Val Ser Glu Asp Glu Asn Asp Thr Ala Ser Arg Thr Pro \*  
                                  1475                      1480                      1485

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:  
 30            (A) LENGTH: 3288 base pairs  
               (B) TYPE: nucleic acid  
               (C) STRANDEDNESS: double  
               (D) TOPOLOGY: linear  
 35            (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

40    ATG GTG ACT GTT ATG CAA AAT AAA ATA TCA TTT TTA TCA GGT ACA TCC 48  
       Met Val Thr Val Met Gln Asn Lys Ile Ser Phe Leu Ser Gly Thr Ser  
       1                      5                      10                      15  
 45    GAA CAG CCC CTG CTT GAC GCC GGT TAT CAA AAC GTA TTT GAT ATC GCA 96  
       Glu Gln Pro Leu Leu Asp Ala Gly Tyr Gln Asn Val Phe Asp Ile Ala  
                                  20                      25                      30  
 50    TCA ATC AGC CGG GCT ACT TTC GTT CAA TCC GTT CCC ACC CTG CCC GTT 144  
       Ser Ile Ser Arg Ala Thr Phe Val Gln Ser Val Pro Thr Leu Pro Val  
                                  35                      40                      45  
 55    AAA GAG GCT CAT ACC GTC TAT CGT CAG GCG CGG CAA CGT GCG GAA AAT 192  
       Lys Glu Ala His Thr Val Tyr Arg Gln Ala Arg Gln Arg Ala Glu Asn  
                                  50                      55                      60  
 60    CTC AAA TCC CTC TAC CGA GCC TGG CAA TTG CGT CAG GAG CCG GTT ATT 240  
       Leu Lys Ser Leu Tyr Arg Ala Trp Gln Leu Arg Gln Glu Pro Val Ile  
                                  65                      70                      75                      80  
 65    AAA GGG CTG GCT AAA CTT AAC CTA CAA TCC AAC GTT TCT GTG CTT CAA 288  
       Lys Gly Leu Ala Lys Leu Asn Leu Gln Ser Asn Val Ser Val Leu Gln  
                                  85                      90                      95  
 65    GAT GCT TTG GTA GAG AAT ATT GGC GGT CAT GGG GAT TTC AGC GAT TTA 336  
       Asp Ala Leu Val Glu Asn Ile Gly Gly Asp Gly Asp Ph Ser Asp Leu



|    | 100                                                                                                                                                       | 105 | 110 |  |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-----|--|
| 5  | ATG AAC CGT GCC AGT CAA TAT GCT GAC GCT GCC TCT ATT CAA TCC CTA 334<br>Met Asn Arg Ala Ser Gln Tyr Ala Asp Ala Ala Ser Ile Gln Ser Leu<br>115 120 125     |     |     |  |
| 10 | TTT TCA CCG GGC CGT TAT GCT TCC GCA CTC TAC AGA GTT GCT AAA GAT 432<br>Phe Ser Pro Gly Arg Tyr Ala Ser Ala Leu Tyr Arg Val Ala Lys Asp<br>130 135 140     |     |     |  |
| 15 | CTG CAT AAA TCA GAT TCC AGT TTG CAT ATT GAT AAT CGC CGC GCT GAT 480<br>Leu His Lys Ser Asp Ser Ser Leu His Ile Asp Asn Arg Arg Ala Asp<br>145 150 155 160 |     |     |  |
| 20 | CTG AAG GAT CTG ATA TTA AGC GAA ACG ACG ATG AAT AAA GAG GTC ACT 528<br>Leu Lys Asp Leu Ile Leu Ser Glu Thr Thr Met Asn Lys Glu Val Thr<br>165 170 175     |     |     |  |
| 25 | TCC CTT GAT ATC TTG TTG GAT GTG CTA CAA AAA GGC GGT AAA GAT ATT 576<br>Ser Leu Asp Ile Leu Leu Asp Val Leu Gln Lys Gly Gly Lys Asp Ile<br>180 185 190     |     |     |  |
| 30 | ACT GAG CTG TCC GGC GCA TTC TTC CCA ATG ACG TTA CCT TAT GAC GAT 624<br>Thr Glu Leu Ser Gly Ala Phe Phe Pro Met Thr Leu Pro Tyr Asp Asp<br>195 200 205     |     |     |  |
| 35 | CAT CTG TCG CAA ATC GAT TCC GCT TTA TCG GCA CAA GCC AGA ACG CTG 672<br>His Leu Ser Gln Ile Asp Ser Ala Leu Ser Ala Gln Ala Arg Thr Leu<br>210 215 220     |     |     |  |
| 40 | AAC GGT GTG TGG AAT ACT TTG ACA GAT ACC ACG GCA CAA GCG GTT TCA 720<br>Asn Gly Val Trp Asn Thr Leu Thr Asp Thr Thr Ala Gln Ala Val Ser<br>225 230 235 240 |     |     |  |
| 45 | GAA CAA ACC AGT AAT ACG AAT ACA CGC AAA CTG TTC GCT GCC CAA GAT 768<br>Glu Gln Thr Ser Asn Thr Asn Thr Arg Lys Leu Phe Ala Ala Gln Asp<br>245 250 255     |     |     |  |
| 50 | GGT AAT CAA GAT ACA TTT TTT TCC GGA AAC ACT TTT TAT TTC AAA GCG 816<br>Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr Phe Tyr Phe Lys Ala<br>260 265 270     |     |     |  |
| 55 | GTG GGA TTC AGC GGG CAA CCT ATG GTT TAC CTG TCA CAG TAC ACC AGC 864<br>Val Gly Phe Ser Gly Gln Pro Met Val Tyr Leu Ser Gln Tyr Thr Ser<br>275 280 285     |     |     |  |
| 60 | GGG AAC GGC ATT GTC GGC GCA CAA TTG ATT GCA GGT AAT CCA GAC CAA 912<br>Gly Asn Gly Ile Val Gly Ala Gln Leu Ile Ala Gly Asn Pro Asp Gln<br>290 295 300     |     |     |  |
| 65 | GCC GCC GCC GCA ATA GTC GCA CCG TTG AAA CTC ACT TGG TCA ATG GCA 960<br>Ala Ala Ala Ala Ile Val Ala Pro Leu Lys Leu Thr Trp Ser Met Ala<br>305 310 315 320 |     |     |  |
| 70 | AAA CAG TGT TAC TAC CTC GTC GCT CCC GAT GGT ACA ACG ATG GGA GAC 1008<br>Lys Gln Cys Tyr Tyr Leu Val Ala Pro Asp Gly Thr Thr Met Gly Asp<br>325 330 335    |     |     |  |
| 75 | GGT AAT GTT CTG ACC GGC TGT TTC TTA AGA GGC AAC AGC CCA ACT AAC 1056<br>Gly Asn Val Leu Thr Gly Cys Phe Leu Arg Gly Asn Ser Pro Thr Asn<br>340 345 350    |     |     |  |
| 80 | CCG GAT AAA GAC GGT ATT TTT GCT CAG GTA GCC AAC AAA TCA GGC AGT 1104<br>Pro Asp Lys Asp Gly Ile Phe Ala Gln Val Ala Asn Lys Ser Gly Ser<br>355 360 365    |     |     |  |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | ACT | CAG | CCT | TTG | CCA | AGC | TTC | CAT | CTG | CCG | GTC | ACA | CTG | GAA | CAC | AGC | 1152 |
|    | Thr | Gln | Pro | Leu | Pro | Ser | Phe | His | Leu | Pro | Val | Thr | Leu | Glu | His | Ser |      |
|    | 370 |     |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |      |
| 5  | GAG | AAT | AAA | GAT | CAG | TAC | TAT | CTG | AAA | ACA | GAG | CAG | GGT | TAT | ATC | ACG | 1200 |
|    | Glu | Asn | Lys | Asp | Gln | Tyr | Tyr | Leu | Lys | Thr | Glu | Gln | Gly | Tyr | Ile | Thr |      |
|    | 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |      |
| 10 | GTA | GAT | AGT | TCC | GGA | CAG | TCA | AAT | TGG | AAA | AAC | GCG | CTG | GTT | ATC | AAT | 1248 |
|    | Val | Asp | Ser | Ser | Gly | Gln | Ser | Asn | Trp | Lys | Asn | Ala | Leu | Val | Ile | Asn |      |
|    |     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |      |
| 15 | GGG | ACA | AAA | GAC | AAG | GGG | CTG | TTA | TTA | ACC | TTT | TGC | AGC | GAT | AGC | TCA | 1296 |
|    | Gly | Thr | Lys | Asp | Lys | Gly | Leu | Leu | Leu | Thr | Phe | Cys | Ser | Asp | Ser | Ser |      |
|    |     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |      |
| 20 | GGC | ACT | CCG | ACA | AAC | CCT | GAT | GAT | GTG | ATT | CCT | CCC | GCT | ATC | AAT | GAT | 1344 |
|    | Gly | Thr | Pro | Thr | Asn | Pro | Asp | Asp | Val | Ile | Pro | Pro | Ala | Ile | Asn | Asp |      |
|    |     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |      |
| 25 | ATT | CCA | TCG | CCG | CCA | GCC | CGC | GAA | ACA | CTG | TCA | CTG | ACG | CCG | GTC | AGT | 1392 |
|    | Ile | Pro | Ser | Pro | Pro | Ala | Arg | Glu | Thr | Leu | Ser | Leu | Thr | Pro | Val | Ser |      |
|    |     |     |     | 450 |     |     | 455 |     |     |     |     | 460 |     |     |     |     |      |
| 30 | TAT | CAA | TTG | ATG | ACC | AAT | CCG | GCA | CCG | ACA | GAA | GAT | GAT | ATT | ACC | AAC | 1440 |
|    | Tyr | Gln | Leu | Met | Thr | Asn | Pro | Ala | Pro | Thr | Glu | Asp | Asp | Ile | Thr | Asn |      |
|    | 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |      |
| 35 | CAT | TAT | GGT | TTT | AAC | GGC | GCT | AGC | TTA | CGG | GCT | TCT | CCA | TTG | TCA | ACC | 1488 |
|    | His | Tyr | Gly | Phe | Asn | Gly | Ala | Ser | Leu | Arg | Ala | Ser | Pro | Leu | Ser | Thr |      |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |      |
| 40 | AGC | GAG | TTG | ACC | AGC | AAA | CTG | AAT | TCT | ATC | GAT | ACT | TTC | TGT | GAG | AAG | 1536 |
|    | Ser | Glu | Leu | Thr | Ser | Lys | Leu | Asn | Ser | Ile | Asp | Thr | Phe | Cys | Glu | Lys |      |
|    |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |      |
| 45 | ACC | CGG | TTA | AGC | TTC | AAT | CAG | TTA | ATG | GAT | TTG | ACC | GCT | CAG | CAA | TCT | 1584 |
|    | Thr | Arg | Leu | Ser | Phe | Asn | Gln | Leu | Met | Asp | Leu | Thr | Ala | Gln | Gln | Ser |      |
|    |     |     |     | 515 |     |     |     | 520 |     |     |     |     | 525 |     |     |     |      |
| 50 | TAC | AGT | CAA | AGC | AGC | ATT | GAT | GCG | AAA | GCA | GCC | AGC | CGC | TAT | GTT | CGT | 1632 |
|    | Tyr | Ser | Gln | Ser | Ser | Ile | Asp | Ala | Lys | Ala | Ala | Ser | Arg | Tyr | Val | Arg |      |
|    |     |     |     | 530 |     |     | 535 |     |     |     |     | 540 |     |     |     |     |      |
| 55 | TTT | GGG | GAA | ACC | ACC | CCA | ACC | CGC | GTC | AAT | GTC | TAC | GGT | GCC | GCT | TAT | 1680 |
|    | Phe | Gly | Glu | Thr | Thr | Pro | Thr | Arg | Val | Asn | Val | Tyr | Gly | Ala | Ala | Tyr |      |
|    | 545 |     |     |     |     | 550 |     |     |     | 555 |     |     |     |     |     | 560 |      |
| 60 | CTG | AAC | AGC | ACA | CTG | GCA | GAC | GCG | GCT | GAT | GGT | CAA | TAT | CTG | TGG | ATT | 1728 |
|    | Leu | Asn | Ser | Thr | Leu | Ala | Asp | Ala | Ala | Asp | Gly | Gln | Tyr | Leu | Trp | Ile |      |
|    |     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |      |
| 65 | CAG | ACT | GAT | GGC | AAG | AGC | CTA | AAT | TTC | ACT | GAC | GAT | ACG | GTA | GTC | GCC | 1776 |
|    | Gln | Thr | Asp | Gly | Lys | Ser | Leu | Asn | Phe | Thr | Asp | Asp | Thr | Val | Val | Ala |      |
|    |     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |      |
| 70 | TTA | GCC | GGT | CGC | GCT | GAA | AAG | CTG | GTA | CGT | TTA | TCA | TCC | CAG | ACC | GGG | 1824 |
|    | Leu | Ala | Gly | Arg | Ala | Glu | Lys | Leu | Val | Arg | Leu | Ser | Ser | Gln | Thr | Gly |      |
|    |     |     |     | 595 |     |     |     | 600 |     |     |     |     | 605 |     |     |     |      |
| 75 | CTA | TCA | TTT | GAA | GAA | TTG | GAC | TGG | CTG | ATT | GCC | AAT | GCC | AGT | CGT | AGT | 1872 |
|    | Leu | Ser | Phe | Glu | Glu | Leu | Asp | Trp | Leu | Ile | Ala | Asn | Ala | Ser | Arg | Ser |      |
|    |     |     |     | 610 |     |     | 615 |     |     |     |     | 620 |     |     |     |     |      |
| 80 | GTG | CCG | GAC | CAC | CAC | GAC | AAA | ATT | GTG | CTG | GAT | AAG | CCG | GTC | CTT | GAA | 1920 |
|    | Val | Pro | Asp | His | His | Asp | Lys | Ile | Val | Leu | Asp | Lys | Pro | Val | Leu | Glu |      |

|    |                                                                 |      |                                                                 |     |     |     |     |  |
|----|-----------------------------------------------------------------|------|-----------------------------------------------------------------|-----|-----|-----|-----|--|
|    | 625                                                             |      | 630                                                             |     | 635 |     | 640 |  |
| 5  | GCA CTG GCA GAG TAT GTC AGC CTA AAA CAG CGC TAT GGG GTT GAT GCC | 1953 | Ala Leu Ala Glu Tyr Val Ser Leu Lys Gln Arg Tyr Gly Leu Asp Ala | 645 | 650 | 655 |     |  |
| 10 | AAT ACC TTT GCG ACC TTC ATT AGT GCA GTA AAT CCT TAT ACG CCA GAT | 2015 | Asn Thr Phe Ala Thr Phe Ile Ser Ala Val Asn Pro Tyr Thr Pro Asp | 660 | 665 | 670 |     |  |
| 15 | CAG ACA CCC AGT TTC TAT GAA ACC GCT TTC CGC TCT GCC GAC GGT AAT | 2064 | Gln Thr Pro Ser Phe Tyr Glu Thr Ala Phe Arg Ser Ala Asp Gly Asn | 675 | 680 | 685 |     |  |
| 20 | CAT GTC ATT GCG CTA GGT ACA GAG GTG AAA TAT GCA GAA AAT GAG CAG | 2112 | His Val Ile Ala Leu Gly Thr Glu Val Lys Tyr Ala Glu Asn Glu Gln | 690 | 695 | 700 |     |  |
| 25 | GAT GAG TTA GCC GCC ATA TGC TGC AAA GCA TTG GGT GTC ACC AGT GAT | 2160 | Asp Glu Leu Ala Ala Ile Cys Cys Lys Ala Leu Gly Val Thr Ser Asp | 705 | 710 | 715 | 720 |  |
| 30 | GAA CTG CTC CGT ATT GGT CGC TAT TGC TTC GGT AAT GCA GGC AGT TTT | 2208 | Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn Ala Gly Ser Phe | 725 | 730 | 735 |     |  |
| 35 | ACC TTG GAT GAA TAT ACC GCC AGT CAG TTG TAT CGC TTC GGC GCC ATT | 2256 | Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg Phe Gly Ala Ile | 740 | 745 | 750 |     |  |
| 40 | CCC CGT TTG TTT GGG CTG ACA TTT GCC CAA GCC GAA ATT TTA TGG CGT | 2304 | Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu Ile Leu Trp Arg | 755 | 760 | 765 |     |  |
| 45 | CTG ATG GAA GGC GGA AAA GAT ATC TTA TTG CAA CAG TTA GGT CAG GCA | 2352 | Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln Leu Gly Gln Ala | 770 | 775 | 780 |     |  |
| 50 | AAA TCC CTG CAA CCA CTG GCT ATT TTA CGC CGT ACC GAG CAG GTG CTG | 2400 | Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr Glu Gln Val Leu | 785 | 790 | 795 | 800 |  |
| 55 | GAT TGG ATG TCG TCC GTA AAT CTA AGT CTG ACT TAT CTG CAA GGG ATG | 2448 | Asp Trp Met Ser Ser Val Asn Leu Ser Leu Thr Tyr Leu Gln Gly Met | 805 | 810 | 815 |     |  |
| 60 | GTA AGT ACG CAA TGG AGC GGT ACC GCC ACC GCT GAG ATG TTC AAT TTC | 2496 | Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu Met Phe Asn Phe | 820 | 825 | 830 |     |  |
| 65 | TTG GAA AAC GTT TGT GAC AGC GTG AAT AGT CAA GCT GCC ACT AAA GAA | 2544 | Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala Ala Thr Lys Glu | 835 | 840 | 845 |     |  |
| 70 | ACA ATG GAT TCG GCG TTA CAG CAG AAA GTG CTG CGG GCG CTA AGC GCC | 2592 | Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg Ala Leu Ser Ala | 850 | 855 | 860 |     |  |
| 75 | GGT TTC GGC ATT AAG AGC AAT GTG ATG GGT ATC GTC ACC TTC TGG CTG | 2640 | Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val Thr Phe Trp Leu | 865 | 870 | 875 | 880 |  |
| 80 | GAG AAA ATC ACA ATC GGT AGT GAT AAT CCT TTT ACA TTG GCA AAC TAC | 2688 | Glu Lys Ile Thr Ile Gly Ser Asp Asn Pro Phe Thr Leu Ala Asn Tyr | 885 | 890 | 895 |     |  |

TGG CAT GAT ATT CAA ACC CTG TTT AGC CAT GAC AAT GCC ACG TTA GAG 2784  
 Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn Ala Thr Leu Glu  
 900 905 910

5 TCC TTA CAA ACC GAC ACT TCT CTG GTA ATT GCT ACT CAG CAA CTT AGC 2784  
 Ser Leu Gln Thr Asp Thr S r Leu Val Ile Ala Thr Gln Gln Leu Ser  
 915 920 925

10 CAG CTA GTG TTA ATT GTG AAA TGG CTG AGC CTG ACC GAG CAG GAT CTG 2832  
 Gln Leu Val Leu Ile Val Lys Trp Leu Ser Leu Thr Glu Gln Asp Leu  
 930 935 940

15 CAA TTA CTG ACA ACC TAT CCC GAA CGT TTA ATC AAC GGC ATC ACG AAT 2880  
 Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn Gly Ile Thr Asn  
 945 950 955 960

20 GTT CCT GTA CCC AAT CCG GAG CTA TTA CTC ACG CTA TCA CGT TTT AAG 2928  
 Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu Ser Arg Phe Lys  
 965 970 975

CAG TGG GAA ACT CAA GTC ACC GTT TCC CGT GAT GAA GCG ATG CGC TGT 2976  
 Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu Ala Met Arg Cys  
 980 985 990

25 TTC GAT CAA TTA AAT GCC AAT GAT ATG ACG ACT GAA AAT GCA GGT TCA 3024  
 Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu Asn Ala Gly Ser  
 995 1000 1005

30 CTG ATC GCC ACA TTG TAT GAG ATG GAT AAA GGT ACG GGA GCG CAA GTT 3072  
 Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr Gly Ala Gln Val  
 1010 1015 1020

35 AAT ACC TTG CTA TTA GGT GAA AAT AAC TGG CCG AAA AGT TTT ACC TCT 3120  
 Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys Ser Phe Thr Ser  
 1025 1030 1035 1040

40 CTC TGG CAA CTT CTG ACC TGG TTA CGC GTC GGG CAA AGA CTG AAT GTC 3168  
 Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln Arg Leu Asn Val  
 1045 1050 1055

GGT AGT ACC ACT CTG GGC AAT CTG TTG TCC ATG ATG CAA GCA GAC CCT 3216  
 Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met Met Gln Ala Asp Pro  
 1060 1065 1070

45 GCT GCC GAG AGT AGC GCT TTA TTG GCA TCA GTA GCC CAA AAC TTA AGT 3264  
 Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala Gln Asn Leu Ser  
 1075 1080 1085

50 GCC GCA ATC AGC AAT CGT CAG TAA 3285  
 Ala Ala Ile Ser Asn Arg Gln ...  
 1090 1095

- (2) INFORMATION FOR SEQ ID NO:34:  
 55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1095 amino acids  
 (B) TYPE: amino acids  
 (C) TOPOLOGY: linear
- 60 (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:  
 Features From To Description  
 254 267 SEQ ID NO:15  
 65 254 492 TcaaIi peptide

5 Met Val Thr Val Met Gln Asn Lys Ile Ser Phe Leu Ser Gly Thr Ser  
 1 5 10 15  
 Glu Gln Pro Leu Leu Asp Ala Gly Tyr Gln Asn Val Phe Asp Ile Ala  
 20 25 30  
 10 Ser Ile Ser Arg Ala Thr Phe Val Gln Ser Val Pro Thr Leu Pro Val  
 35 40 45  
 Lys Glu Ala His Thr Val Tyr Arg Gln Ala Arg Gln Arg Ala Glu Asn  
 50 55 60  
 15 Leu Lys Ser Leu Tyr Arg Ala Trp Gln Leu Arg Gln Glu Pro Val Ile  
 65 70 75 80  
 Lys Gly Leu Ala Lys Leu Asn Leu Gln Ser Asn Val Ser Val Leu Gln  
 85 90 95  
 20 Asp Ala Leu Val Glu Asn Ile Gly Gly Asp Gly Asp Phe Ser Asp Leu  
 100 105 110  
 25 Met Asn Arg Ala Ser Gln Tyr Ala Asp Ala Ala Ser Ile Gln Ser Leu  
 115 120 125  
 Phe Ser Pro Gly Arg Tyr Ala Ser Ala Leu Tyr Arg Val Ala Lys Asp  
 130 135 140  
 30 Leu His Lys Ser Asp Ser Ser Leu His Ile Asp Asn Arg Arg Ala Asp  
 145 150 155 160  
 Leu Lys Asp Leu Ile Leu Ser Glu Thr Thr Met Asn Lys Glu Val Thr  
 165 170 175  
 35 Ser Leu Asp Ile Leu Leu Asp Val Leu Gln Lys Gly Gly Lys Asp Ile  
 180 185 190  
 40 Thr Glu Leu Ser Gly Ala Phe Phe Pro Met Thr Leu Pro Tyr Asp Asp  
 195 200 205  
 His Leu Ser Gln Ile Asp Ser Ala Leu Ser Ala Gln Ala Arg Thr Leu  
 210 215 220  
 45 Asn Gly Val Trp Asn Thr Leu Thr Asp Thr Thr Ala Gln Ala Val Ser  
 225 230 235 240  
 Glu Gln Thr Ser Asn Thr Asn Thr Arg Lys Leu Phe Ala Ala Gln Asp  
 245 250 255  
 50 Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr Phe Tyr Phe Lys Ala  
 260 265 270  
 55 Val Gly Phe Ser Gly Gln Pro Met Val Tyr Leu Ser Gln Tyr Thr Ser  
 275 280 285  
 Gly Asn Gly Ile Val Gly Ala Gln Leu Ile Ala Gly Asn Pro Asp Gln  
 290 295 300  
 60 Ala Ala Ala Ala Ile Val Ala Pro Leu Lys Leu Thr Trp Ser Met Ala  
 305 310 315 320  
 Lys Gln Cys Tyr Tyr Leu Val Ala Pro Asp Gly Thr Thr Met Gly Asp  
 325 330 335  
 65 Gly Asn Val Leu Thr Gly Cys Phe Leu Arg Gly Asn Ser Pro Thr Asn

|    | 340                                                                                 | 345 | 350 |
|----|-------------------------------------------------------------------------------------|-----|-----|
| 5  | Pro Asp Lys Asp Gly Ile Phe Ala Gln Val Ala Asn Lys Ser Gly Ser<br>355 360 365      |     |     |
|    | Thr Gln Pro Leu Pro Ser Phe His Leu Pro Val Thr Leu Glu His Ser<br>370 375 380      |     |     |
| 10 | Glu Asn Lys Asp Gln Tyr Tyr Leu Lys Thr Glu Gln Gly Tyr Ile Thr<br>385 390 395 400  |     |     |
|    | Val Asp Ser Ser Gly Gln Ser Asn Trp Lys Asn Ala Leu Val Ile Asn<br>405 410 415      |     |     |
| 15 | Gly Thr Lys Asp Lys Gly Leu Leu Leu Thr Phe Cys Ser Asp Ser Ser<br>420 425 430      |     |     |
|    | Gly Thr Pro Thr Asn Pro Asp Asp Val Ile Pro Pro Ala Ile Asn Asp<br>435 440 445      |     |     |
| 20 | Ile Pro Ser Pro Pro Ala Arg Glu Thr Leu Ser Leu Thr Pro Val Ser<br>450 455 460      |     |     |
| 25 | Tyr Gln Leu Met Thr Asn Pro Ala Pro Thr Glu Asp Asp Ile Thr Asn<br>465 470 475 480  |     |     |
|    | His Tyr Gly Phe Asn Gly Ala Ser Leu Arg Ala Ser Pro Leu Ser Thr<br>485 490 W4 » 495 |     |     |
| 30 | ser Glu Leu Thr Ser Lys Leu Asn Ser Ile Asp Thr Phe Cys Glu Lys<br>500 505 510      |     |     |
|    | Thr Arg Leu Ser Phe Asn Gln Leu Met Asp Leu Thr Ala Gln Gln Ser<br>515 520 525      |     |     |
| 35 | Tyr Ser Gln Ser Ser Ile Asp Ala Lys Ala Ala Ser Arg Tyr Val Arg<br>530 535 540      |     |     |
| 40 | Phe Gly Glu Thr Thr Pro Thr Arg Val Asn Val Tyr Gly Ala Ala Tyr<br>545 550 555 560  |     |     |
|    | Leu Asn Ser Thr Leu Ala Asp Ala Ala Asp Gly Gln Tyr Leu Trp Ile<br>565 570 575      |     |     |
| 45 | Gln Thr Asp Gly Lys Ser Leu Asn Phe Thr Asp Asp Thr Val Val Ala<br>580 585 590      |     |     |
|    | Leu Ala Gly Arg Ala Glu Lys Leu Val Arg Leu Ser Ser Gln Thr Gly<br>595 600 605      |     |     |
| 50 | Leu Ser Phe Glu Glu Leu Asp Trp Leu Ile Ala Asn Ala Ser Arg Ser<br>610 615 620      |     |     |
| 55 | Val Pro Asp His His Asp Lys Ile Val Leu Asp Lys Pro Val Leu Glu<br>625 630 635 640  |     |     |
|    | Ala Leu Ala Glu Tyr Val Ser Leu Lys Gln Arg Tyr Gly Leu Asp Ala<br>645 650 655      |     |     |
| 60 | Asn Thr Phe Ala Thr Phe Ile Ser Ala Val Asn Pro Tyr Thr Pro Asp<br>660 665 670      |     |     |
|    | Gln Thr Pro Ser Phe Tyr Glu Thr Ala Phe Arg Ser Ala Asp Gly Asn<br>675 680 685      |     |     |
| 65 | His Val Ile Ala Leu Gly Thr Glu Val Lys Tyr Ala Glu Asn Glu Gln                     |     |     |

|    |      |     |     |     |     |      |      |      |     |     |      |      |      |     |     |      |     |  |  |
|----|------|-----|-----|-----|-----|------|------|------|-----|-----|------|------|------|-----|-----|------|-----|--|--|
|    | 690  |     | 695 |     | 700 |      |      |      |     |     |      |      |      |     |     |      |     |  |  |
|    | Asp  | Glu | Leu | Ala | Ala | Ile  | Cys  | Cys  | Lys | Ala | Leu  | Gly  | Val  | Thr | Ser | Asp  |     |  |  |
|    | 705  |     |     |     |     | 710  |      |      |     |     | 715  |      |      |     |     | 720  |     |  |  |
| 5  |      | Glu | Leu | Leu | Arg | Ile  | Gly  | Arg  | Tyr | Cys | Phe  | Gly  | Asn  | Ala | Gly | Ser  | Phe |  |  |
|    |      |     |     |     | 725 |      |      |      |     |     | 730  |      |      |     |     | 735  |     |  |  |
|    | Thr  | Leu | Asp | Glu | Tyr | Thr  | Ala  | Ser  | Gln | Leu | Tyr  | Arg  | Phe  | Gly | Ala | Ile  |     |  |  |
| 10 |      |     |     | 740 |     |      |      |      | 745 |     |      |      |      | 750 |     |      |     |  |  |
|    | Pro  | Arg | Leu | Phe | Gly | Leu  | Thr  | Phe  | Ala | Gln | Ala  | Glu  | Ile  | Leu | Trp | Arg  |     |  |  |
|    |      |     | 755 |     |     |      |      | 760  |     |     |      |      | 765  |     |     |      |     |  |  |
| 15 | Leu  | Met | Glu | Gly | Gly | Lys  | Asp  | Ile  | Leu | Leu | Gln  | Gln  | Leu  | Gly | Gln | Ala  |     |  |  |
|    | 770  |     |     |     |     |      | 775  |      |     |     |      | 780  |      |     |     |      |     |  |  |
|    | Lys  | Ser | Leu | Gln | Pro | Leu  | Ala  | Ile  | Leu | Arg | Arg  | Thr  | Glu  | Gln | Val | Leu  |     |  |  |
|    | 785  |     |     |     |     | 790  |      |      |     |     | 795  |      |      |     |     | 800  |     |  |  |
| 20 |      | Asp | Trp | Met | Ser | Ser  | Val  | Asn  | Leu | Ser | Leu  | Thr  | Tyr  | Leu | Gln | Gly  | Met |  |  |
|    |      |     |     |     |     | 805  |      |      |     |     | 810  |      |      |     |     | 815  |     |  |  |
|    | Val  | Ser | Thr | Gln | Trp | Ser  | Gly  | Thr  | Ala | Thr | Ala  | Glu  | Met  | Phe | Asn | Phe  |     |  |  |
| 25 |      |     |     | 820 |     |      |      |      | 825 |     |      |      |      | 830 |     |      |     |  |  |
|    | Leu  | Glu | Asn | Val | Cys | Asp  | Ser  | Val  | Asn | Ser | Gln  | Ala  | Ala  | Thr | Lys | Glu  |     |  |  |
|    |      |     | 835 |     |     |      |      | 840  |     |     |      | 845  |      |     |     |      |     |  |  |
| 30 | Thr  | Met | Asp | Ser | Ala | Leu  | Gln  | Gln  | Lys | Val | Leu  | Arg  | Ala  | Leu | Ser | Ala  |     |  |  |
|    | 850  |     |     |     |     |      | 855  |      |     |     |      | 860  |      |     |     |      |     |  |  |
|    | Gly  | Phe | Gly | Ile | Lys | Ser  | Asn  | Val  | Met | Gly | Ile  | Val  | Thr  | Phe | Trp | Leu  |     |  |  |
|    | 865  |     |     |     |     | 870  |      |      |     |     | 875  |      |      |     |     | 880  |     |  |  |
| 35 |      | Glu | Lys | Ile | Thr | Ile  | Gly  | Ser  | Asp | Asn | Pro  | Phe  | Thr  | Leu | Ala | Asn  | Tyr |  |  |
|    |      |     |     |     | 885 |      |      |      |     | 890 |      |      |      |     |     | 895  |     |  |  |
|    | Trp  | His | Asp | Ile | Gln | Thr  | Leu  | Phe  | Ser | His | Asp  | Asn  | Ala  | Thr | Leu | Glu  |     |  |  |
| 40 |      |     |     | 900 |     |      |      |      | 905 |     |      |      |      | 910 |     |      |     |  |  |
|    | Ser  | Leu | Gln | Thr | Asp | Thr  | Ser  | Leu  | Val | Ile | Ala  | Thr  | Gln  | Gln | Leu | Ser  |     |  |  |
|    |      |     | 915 |     |     |      |      | 920  |     |     |      |      | 925  |     |     |      |     |  |  |
| 45 | Gln  | Leu | Val | Leu | Ile | Val  | Lys  | Trp  | Leu | Ser | Leu  | Thr  | Glu  | Gln | Asp | Leu  |     |  |  |
|    | 930  |     |     |     |     |      | 935  |      |     |     |      | 940  |      |     |     |      |     |  |  |
|    | Gln  | Leu | Leu | Thr | Thr | Tyr  | Pro  | Glu  | Arg | Leu | Ile  | Asn  | Gly  | Ile | Thr | Asn  |     |  |  |
|    | 945  |     |     |     |     | 950  |      |      |     |     | 955  |      |      |     |     | 960  |     |  |  |
| 50 |      | Val | Pro | Val | Pro | Asn  | Pro  | Glu  | Leu | Leu | Leu  | Thr  | Leu  | Ser | Arg | Phe  | Lys |  |  |
|    |      |     |     |     | 965 |      |      |      |     |     | 970  |      |      |     |     | 975  |     |  |  |
|    | Gln  | Trp | Glu | Thr | Gln | Val  | Thr  | Val  | Ser | Arg | Asp  | Glu  | Ala  | Met | Arg | Cys  |     |  |  |
| 55 |      |     |     | 980 |     |      |      |      | 985 |     |      |      |      | 990 |     |      |     |  |  |
|    | Phe  | Asp | Gln | Leu | Asn | Ala  | Asn  | Asp  | Met | Thr | Thr  | Glu  | Asn  | Ala | Gly | Ser  |     |  |  |
|    |      |     | 995 |     |     |      |      | 1000 |     |     |      |      | 1005 |     |     |      |     |  |  |
| 60 | Leu  | Ile | Ala | Thr | Leu | Tyr  | Glu  | Met  | Asp | Lys | Gly  | Thr  | Gly  | Ala | Gln | Val  |     |  |  |
|    | 1010 |     |     |     |     |      | 1015 |      |     |     |      | 1020 |      |     |     |      |     |  |  |
|    | Asn  | Thr | Leu | Leu | Leu | Gly  | Glu  | Asn  | Asn | Trp | Pro  | Lys  | Ser  | Phe | Thr | Ser  |     |  |  |
|    | 1025 |     |     |     |     | 1030 |      |      |     |     | 1035 |      |      |     |     | 1040 |     |  |  |
| 65 |      | Leu | Trp | Gln | Leu | Leu  | Thr  | Trp  | Leu | Arg | Val  | Gly  | Gln  | Arg | Leu | Asn  | Val |  |  |

[illegible]

(2) INFORMATION FOR SEQ ID NO:35  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 603 amino acids  
 (B) TYPE: amino acid  
 (C) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

[illegible]



Val Thr Ser Asp Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn  
 225 230 235 240  
 5 Ala Gly Arg Phe Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg  
 245 250 255  
 Phe Gly Ala Ile Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu  
 260 265 270  
 10 Ile Leu Trp Arg Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln  
 275 280 285  
 Xxx Gly Gln Ala Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr  
 290 295 300  
 15 Glu Gln Val Leu Asp Trp Met Ser Pro Val Asn Leu Ser Leu Thr Tyr  
 305 310 315 320  
 20 Leu Gln Gly Met Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu  
 325 330 335  
 Met Phe Asn Phe Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala  
 340 345 350  
 25 Xxx Thr Lys Glu Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg  
 355 360 365  
 30 Ala Leu Ser Ala Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val  
 370 375 380  
 Thr Phe Trp Leu Glu Lys Ile Thr Ile Gly Arg Asp Asn Pro Phe Thr  
 385 390 395 400  
 35 Leu Ala Asn Tyr Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn  
 405 410 415  
 Ala Thr Leu Glu Ser Leu Gln Thr Asp Thr Ser Leu Val Ile Ala Thr  
 420 425 430  
 40 Gln Gln Leu Ser Gln Leu Val Leu Ile Val Lys Trp Val Ser Leu Thr  
 435 440 445  
 Glu Gln Asp Leu Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn  
 450 455 460  
 45 Gly Ile Thr Asn Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu  
 465 470 475 480  
 50 Ser Arg Phe Lys Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu  
 485 490 495  
 Ala Met Arg Cys Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu  
 500 505 510  
 55 Asn Ala Gly Ser Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr  
 515 520 525  
 Gly Ala Gln Val Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys  
 530 535 540  
 60 Ser Phe Thr Ser Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln  
 545 550 555 560  
 65 Arg Leu Asn Val Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met M c  
 565 570 575

Gln Ala Asp Pro Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala  
530 585 590

5 Gln Asn Leu Ser Ala Ala Ile Ser Asn Arg Gln \*  
595 600

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 2557 base pairs  
(B) TYPE: nucleic acid  
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GAATTCGGCT TCGGTTTAAT ATTGATGATG TCTCGCTCTT CCGCCTGCTT AAAATTACCG 60  
20 ACCATGATAA TAAAGATGGA AAAATTAATA ATAACCTAAA GAATCTTTCC AATTTATATA 120  
TTGGAAAATT ACTGGCAGAT ATTCATCAAT TAACCATTGA TGAAGTGGAT TTATTACTGA 180  
TTGCCGTAGG TGAAGGAAAA ACTAATTTAT CCGCTATCAG TGATAAGCAA TTGGCTACCC 240  
TGATCAGAAA ACTCAATACT ATTACCAGCT GGCTACATAC ACAGAAGTGG AGTGATTCC 300  
AGCTATTTAT CATGACCTCC ACCAGCTATA ACAAACGCT AACGCCTGAA ATTAAGAATT 360  
25 TGCTGGATAC CGTCTACCAC GGTTTACAAG GTTTTGATAA AGACAAAGCA GATTTGCTAC 420  
ATGTCATGGC GCCCTATATT GCGGCCACCT TGCAATTATC ATCGGAAAAT GTCGCCCCACT 480  
CGGTACTCCT TTGGGCAGAT AAGTTACAGC CCGGCGACGG CGCAATGACA GCAGAGGGAN 540  
TCTGGGACTG GTTGAATACT AAGTATACGC CGGGTTTCATC GGAAGCCGTA GAAACGCAGG 600  
AACATATCGT TCAGTATTGT CAGGCTCTGG CACAATTGGA AATGGTTTAC CATTCACCG 660  
30 GCATCAACGA AAACGCCCTC CGTCTATTTG TGACAAAACC AGAGATGTTT GCGCCTGCAA 720  
CTGGAGCAGC GCCCGCGCAT GATGCCCTTT CACTGATTAT GCTGACACGT TTGCGGATT 780  
GGGTGAACGC ACTAGGCGAA AAAGCGTCCT CGGTGCTAGC GGCATTTGAA GCTAACTCGT 840  
TAACGGCAGA ACAACTGGCT GATGCCATGA ATCTTGATGC TAATTTGCTG TTGCAAGCCA 900  
GTATTCAAGC ACAAATCAT CAACATCTTC CCCCAGTAAC TCCAGAAAAT GCGTTCTCCT 960  
35 GTTGGACATC TATCAATACT ATCCTGCAAT GGGTTAATGT CGCACAACAA TTGAAATGTC 1020  
GCCCCACAGG GCGTTTCCGC TTTGGTCCGG CTGGATTATA TTCAATCAAT GAAAGAGACA 1080  
CCGACCTATG CCCAGTGGGA AAACGCGGCA GCGGTATTAA CCGCCGGGTT GAATTCACAA 1140  
ACAGGCTAAT ACATTACAAC GCTTTTCTGG ATGAATCTCG CAGTGCCGCA TTAAGCACCT 1200  
ACTATATCCG TCAAGTCGCC AAGGCAGCGG CGGCTATTAA AAGCCGTGAT GACTTGTATC 1260  
40 AATACTTACT GATTGATAAT CAGGTTTCTG CGGCAATAAA AACCACCCGG ATCGCCGAAG 1320  
CCATTGCCAG TATTCAACTG TACGTCAACC GGGCATTGGA AAATGTGGAA GAAAATGCCA 1380  
ATTCGGGGGT TATCAGCCGC CAATTCTTTA TCGACTGGGA CAAATACAAT AAACGCTACA 1440  
GCACTTGGGC GGGTGTTTCT CAATTAGTTT ACTACCCGGA AACTATATT GATCCGACCA 1500  
TGCGTATCGG ACAAACCAA ATGATGGACG CATTACTGCA ATCCGTCAGC CAAAGCCAA 1560  
45 TAAACGCCGA TACCGTCGAA GATGCCTTTA TGTCTTATCT GACATCGTTT GAACAAGTGG 1620  
CTAATCTTAA AGTTATTAGC GCATATCAG ATAATATTAA TAACGATCAA GGGCTGACCT 1680  
ATTTTATCGG ACTCAGTGAA ACTGATGCCG GTGAATATTA TTGGCGCAGT GTCGATCACA 1740  
GTAAATTCAA CGACGGTAAA TTCGCGGCTA ATGCCTGGAG TGAATGGCAT AAAATTGATT 1800  
GTCCAATTAA CCCTTATAAA AGCACTATCC GTCCAGTGAT ATATAAATCC CGCCTGTATC 1860

```

TGCTCTGGTT GGAACAAAAG GAGATCACCA AACAGACAGG AAATAGTAAA GATGGCTATC 1920
AAACTGAAAC GGATTATCGT TATGAACTAA AATTGGCGCA TATCCGCTAT GATGGCACTT 1980
GGAATACGCC AATCACCTTT GATGTCAATA AAAAAATATC CGAGCTAAAA CTGGAAAAAA 2040
ATAGAGCGCC CGGACTCTAT TGTGCCGGTT ATCAAGGTGA AGATACGTTG CTGGTGATGT 2100
5 TTTATAACCA ACAAGACACA CTAGATAGTT ATAAAAACGC TTCAATGCAA GGACTATATA 2160
TCTTTGCTGA TATGGCATCC AAAGATATGA CCCGAGAACA GAGCAATGTT TATCGGGATA 2220
ATAGCTATCA ACAATTTGAT ACCAATAATG TCAGAAGAGT GAATAACCGC TATGCAGAGG 2280
ATTATGAGAT TCCTTCTTCG GTAAGTAGCC GTAAAGACTA TGGTTGGGGA GATTATTACC 2340
TCAGCATGGT ATATAACGGA GATATTCCAA CTATCAATTA CAAAGCCGCA TCAAGTGATT 2400
10 TAAAAATTTA TATTTACCA AAATTAAGAA TTATTCATAA TGGATATGAA GGACAGAAGC 2460
GCAATCAATG CAATTTGATG AATAAATATG GCAAAC TAGG TGATAAATTT ATTGTGTATA 2520
CCAGCCTGGG CGTTAATCCG AATAATAAGC CGAATTC 2557

```

- 15 (2) INFORMATION FOR SEQ ID NO:37:
- (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 845 amino acids
    - (B) TYPE: amino acids
    - (C) TOPOLOGY: linear
  - 20 (ii) MOLECULE TYPE: protein (partial)
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```

25 Ala Phe Asn Ile Asp Asp Val Ser Leu Phe Arg Leu Leu Lys Ile Thr
 1 5 10 15
30 Asp His Asp Asn Lys Asp Gly Lys Ile Lys Asn Asn Leu Lys Asn Leu
 20 25 30
 Ser Asn Leu Tyr Ile Gly Lys Leu Leu Ala Asp Ile His Gln Leu Thr
 35 40 45
35 Ile Asp Glu Leu Asp Leu Leu Leu Ile Ala Val Gly Glu Gly Lys Thr
 50 55 60
 Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Thr Leu Ile Arg Lys
 65 70 75 80
40 Leu Asn Thr Ile Thr Ser Trp Leu His Thr Gln Lys Trp Ser Val Phe
 85 90 95
 Gln Leu Phe Ile Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr Pro
 100 105 110
 Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly Phe
 115 120 125
50 Asp Lys Asp Lys Ala Asp Leu Leu His Val Met Ala Pro Tyr Ile Ala
 130 135 140
 Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu Leu
 145 150 155 160
55 Trp Ala Asp Lys Leu Gln Pro Gly Asp Gly Ala Met Thr Ala Glu Gly
 165 170 175
 Phe Trp Asp Trp Leu Asn Thr Lys Tyr Thr Pro Gly Ser Ser Glu Ala

```

|    | 180                             | 185                         | 190             |
|----|---------------------------------|-----------------------------|-----------------|
|    | Val Glu Thr Gln Glu His Ile     | Val Gln Tyr Cys Gln Ala     | Leu Ala Gln     |
|    | 195                             | 200                         | 205             |
| 5  | Leu Glu Met Val Tyr His Ser     | Thr Gly Ile Asn Glu         | Asn Ala Phe Arg |
|    | 210                             | 215                         | 220             |
| 10 | Leu Phe Val Thr Lys Pro Glu Met | Phe Gly Ala Ala Thr Gly Ala | Ala             |
|    | 225                             | 230                         | 235             |
|    | Pro Ala His Asp Ala Leu Ser Leu | Ile Met Leu Thr Arg Phe Ala | Asp             |
|    | 245                             | 250                         | 255             |
| 15 | Trp Val Asn Ala Leu Gly Glu Lys | Ala Ser Ser Val Leu Ala Ala | Phe             |
|    | 260                             | 265                         | 270             |
|    | Glu Ala Asn Ser Leu Thr Ala Glu | Gln Leu Ala Asp Ala Met Asn | Leu             |
|    | 275                             | 280                         | 285             |
| 20 | Asp Ala Asn Leu Leu Leu Gln Ala | Ser Ile Gln Ala Gln Asn His | Gln             |
|    | 290                             | 295                         | 300             |
| 25 | His Leu Pro Pro Val Thr Pro Glu | Asn Ala Phe Ser Cys Trp Thr | Ser             |
|    | 305                             | 310                         | 315             |
|    | Ile Asn Thr Ile Leu Gln Trp Val | Asn Val Ala Gln Gln Leu Lys | Cys             |
|    | 325                             | 330                         | 335             |
| 30 | Arg Pro Thr Gly Arg Phe Arg Phe | Gly Arg Ala Gly Leu Tyr Ser | Ile             |
|    | 340                             | 345                         | 350             |
|    | Asn Glu Arg Asp Thr Asp Leu Cys | Pro Val Gly Lys Arg Gly Arg | Arg             |
|    | 355                             | 360                         | 365             |
| 35 | Ile Asn Arg Arg Val Glu Phe Asn | Asn Arg Leu Ile His Tyr Asn | Ala             |
|    | 370                             | 375                         | 380             |
| 40 | Phe Leu Asp Glu Ser Arg Ser Ala | Ala Leu Ser Thr Tyr Tyr Ile | Arg             |
|    | 385                             | 390                         | 395             |
|    | Gln Val Ala Lys Ala Ala Ala Ala | Ile Lys Ser Arg Asp Asp Leu | Tyr             |
|    | 405                             | 410                         | 415             |
| 45 | Gln Tyr Leu Leu Ile Asp Asn Gln | Val Ser Ala Ala Ile Lys Thr | Thr             |
|    | 420                             | 425                         | 430             |
|    | Arg Ile Ala Glu Ala Ile Ala Ser | Ile Gln Leu Tyr Val Asn Arg | Ala             |
|    | 435                             | 440                         | 445             |
| 50 | Leu Glu Asn Val Glu Glu Asn Ala | Asn Ser Gly Val Ile Ser Arg | Gln             |
|    | 450                             | 455                         | 460             |
| 55 | Phe Phe Ile Asp Trp Asp Lys Tyr | Asn Lys Arg Tyr Ser Thr Trp | Ala             |
|    | 465                             | 470                         | 475             |
|    | Gly Val Ser Gln Leu Val Tyr Tyr | Pro Glu Asn Tyr Ile Asp Pro | Thr             |
|    | 485                             | 490                         | 495             |
| 60 | Met Arg Ile Gly Gln Thr Lys Met | Met Asp Ala Leu Leu Gln Ser | Val             |
|    | 500                             | 505                         | 510             |
|    | Ser Gln Ser Gln Leu Asn Ala Asp | Thr Val Glu Asp Ala Phe Met | Ser             |
|    | 515                             | 520                         | 525             |
| 65 | Tyr Leu Thr Ser Phe Glu Gln Val | Ala Asn Leu Lys Val Ile Ser | Ala             |

|    | 530                 | 535                         | 540                             |
|----|---------------------|-----------------------------|---------------------------------|
|    | Tyr His Asp Asn Ile | Asn Asn Asp Gln Gly         | Leu Thr Tyr Phe Ile Gly         |
|    | 545                 | 550                         | 555 560                         |
| 5  | Leu Ser Glu Thr     | Asp Ala Gly Glu Tyr         | Tyr Trp Arg Ser Val Asp His     |
|    |                     | 565                         | 570 575                         |
| 10 | Ser Lys Phe Asn Asp | Gly Lys Phe Ala Ala Asn Ala | Trp Ser Glu Trp                 |
|    |                     | 580                         | 585 590                         |
|    | His Lys Ile Asp Cys | Pro Ile Asn Pro Tyr Lys     | Ser Thr Ile Arg Pro             |
|    |                     | 595                         | 600 605                         |
| 15 | Val Ile Tyr Lys Ser | Arg Leu Tyr Leu Leu Trp     | Leu Glu Gln Lys Glu             |
|    |                     | 610 615                     | 620                             |
|    | Ile Thr Lys Gln Thr | Gly Asn Ser Lys Asp         | Gly Tyr Gln Thr Glu Thr         |
|    |                     | 630                         | 635 640                         |
| 20 | Asp Tyr Arg Tyr     | Glu Leu Lys Leu Ala His     | Ile Arg Tyr Asp Gly Thr         |
|    |                     | 645                         | 650 655                         |
| 25 | Trp Asn Thr         | Pro Ile Thr Phe Asp         | Val Asn Lys Lys Ile Ser Glu Leu |
|    |                     | 660                         | 665 670                         |
|    | Lys Leu Glu Lys     | Asn Arg Ala Pro Gly         | Leu Tyr Cys Ala Gly Tyr Gln     |
|    |                     | 675                         | 680 685                         |
| 30 | Gly Glu Asp Thr     | Leu Leu Val Met Phe Tyr     | Asn Gln Gln Asp Thr Leu         |
|    |                     | 690 695                     | 700                             |
|    | Asp Ser Tyr Lys     | Asn Ala Ser Met Gln Gly     | Leu Tyr Ile Phe Ala Asp         |
|    |                     | 710                         | 715 720                         |
| 35 | Met Ala Ser Lys     | Asp Met Thr Pro Glu         | Gln Ser Asn Val Tyr Arg Asp     |
|    |                     | 725                         | 730 735                         |
| 40 | Asn Ser Tyr Gln     | Gln Phe Asp Thr Asn Asn     | Val Arg Arg Val Asn Asn         |
|    |                     | 740                         | 745 750                         |
|    | Arg Tyr Ala Glu     | Asp Tyr Glu Ile Pro         | Ser Ser Val Ser Ser Arg Lys     |
|    |                     | 755                         | 760 765                         |
| 45 | Asp Tyr Gly Trp     | Gly Asp Tyr Tyr Leu Ser     | Met Val Tyr Asn Gly Asp         |
|    |                     | 770 775                     | 780                             |
|    | Ile Pro Thr Ile     | Asn Tyr Lys Ala Ala Ser     | Ser Asp Leu Lys Ile Tyr         |
|    |                     | 785 790                     | 795 800                         |
| 50 | Ile Ser Pro Lys     | Leu Arg Ile Ile His         | Asn Gly Tyr Glu Gly Gln Lys     |
|    |                     | 805                         | 810 815                         |
| 55 | Arg Asn Gln Cys     | Asn Leu Met Asn Lys Tyr     | Gly Lys Leu Gly Asp Lys         |
|    |                     | 820                         | 825 830                         |
|    | Phe Ile Val Tyr     | Thr Ser Leu Gly Val         | Asn Pro Asn Asn                 |
|    |                     | 835                         | 840 845                         |

60

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

65 (A) LENGTH: 16 amino acids  
(B) TYPE: amino acid

(C) STRANDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly  
1 5 10 15  
Lys

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 amino acids  
(B) TYPE: amino acid  
(C) STRANDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Gly Thr Ala Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala  
1 5 10 15  
Ile Ser Pro Ala Lys  
20

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(C) STRANDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Ala Asn Ser Leu Tyr Ala Leu Phe Leu Pro Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 amino acids

- (B) TYPE: amino acid  
(C) STRANDNESS: single  
(D) TOPOLOGY: linear
- 5 (ii) MOLECULAR TYPE: protein  
(v) FRAGMENT TYPE: N-terminal  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
- 10 Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln  
1 5 10
- 15 (2) INFORMATION FOR SEQ ID NO:42:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
20 (C) STRANDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULAR TYPE: protein  
25 (v) FRAGMENT TYPE: N-terminal  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
- 30 Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Glu Val Tyr  
1 5 10 15  
Ala Gly Leu Glu
- 35 (2) INFORMATION FOR SEQ ID NO:43:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
40 (B) TYPE: amino acid  
(C) STRANDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULAR TYPE: protein  
45 (v) FRAGMENT TYPE: N-terminal  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:
- 50 Ile Arg Glu Asp Tyr Pro Ala Ser Leu Gly Lys  
1 5 10
- 55 (2) INFORMATION FOR SEQ ID NO:44:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 16 amino acids  
60 (B) TYPE: amino acid  
(C) STRANDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

10 Asp Asp Ser Gly Asp Asp Asp Lys Val Thr Asn Thr Asp Ile His  
1 5 10 15  
Arg

15 (2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

20 (C) STRANDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

25 (v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

30 Asp Val Xaa Gly Ser Glu Lys Ala Asn Glu Lys Leu Lys  
1 5 10

(2) INFORMATION FOR SEQ ID NO:46:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7551 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

40 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46 (tcdA):

45 ATG AAC GAG TCT GTA AAA GAG ATA CCT GAT GTA TTA AAA AGC CAG TGT 48  
Met Asn Glu Ser Val Lys Glu Ile Pro Asp Val Leu Lys Ser Gln Cys  
1 5 10 15

50 GGT TTT AAT TGT CTG ACA GAT ATT AGC CAC AGC TCT TTT AAT GAA TTT 96  
Gly Phe Asn Cys Leu Thr Asp Ile Ser His Ser Ser Phe Asn Glu Phe  
20 25 30

55 CGC CAG CAA GTA TCT GAG CAC CTC TCC TGG TCC GAA ACA CAC GAC TTA 144  
Arg Gln Gln Val Ser Glu His Leu Ser Trp Ser Glu Thr His Asp Leu  
35 40 45

60 TAT CAT GAT GCA CAA CAG GCA CAA AAG GAT AAT CGC CTG TAT GAA GCG 192  
Tyr His Asp Ala Gln Gln Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala  
50 55 60

65 CGT ATT CTC AAA CGC GCC AAT CCC CAA TTA CAA AAT GCG GTG CAT CTT 240  
Arg Ile Leu Lys Arg Ala Asn Pro Gln Leu Gln Asn Ala Val His Leu  
65 70 75 80



|    |                                                                 |      |
|----|-----------------------------------------------------------------|------|
|    | GCC ATT CTC GCT CCC AAT GCT GAA CTG ATA GGC TAT AAC AAT CAA TTT | 288  |
|    | Ala Ile Leu Ala Pro Asn Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe |      |
|    | 85 90 95                                                        |      |
| 5  | AGC GGT AGA GCC AGT CAA TAT GTT GCG CCG GGT ACC GTT TCT TCC ATG | 336  |
|    | Ser Gly Arg Ala Ser Gln Tyr Val Ala Pro Gly Thr Val Ser Ser Met |      |
|    | 100 105 110                                                     |      |
| 10 | TTC TCC CCC GCC GCT TAT TTG ACT GAA CTT TAT CGT GAA GCA CGC AAT | 384  |
|    | Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn |      |
|    | 115 120 125                                                     |      |
| 15 | TTA CAC GCA AGT GAC TCC GTT TAT TAT CTG GAT ACC CGC CGC CCA GAT | 432  |
|    | Leu His Ala Ser Asp Ser Val Tyr Tyr Leu Asp Thr Arg Arg Pro Asp |      |
|    | 130 135 140                                                     |      |
| 20 | CTC AAA TCA ATG GCG CTC AGT CAG CAA AAT ATG GAT ATA GAA TTA TCC | 480  |
|    | Leu Lys Ser Met Ala Leu Ser Gln Gln Asn Met Asp Ile Glu Leu Ser |      |
|    | 145 150 155 160                                                 |      |
|    | ACA CTC TCT TTG TCC AAT GAG CTG TTA TTG GAA AGC ATT AAA ACT GAA | 528  |
|    | Thr Leu Ser Leu Ser Asn Glu Leu Leu Glu Ser Ile Lys Thr Thr Glu |      |
|    | 165 170 175                                                     |      |
| 25 | TCT AAA CTG GAA AAC TAT ACT AAA GTG ATG GAA ATG CTC TCC ACT TTC | 576  |
|    | Ser Lys Leu Glu Asn Tyr Thr Lys Val Met Glu Met Leu Ser Thr Phe |      |
|    | 180 185 190                                                     |      |
| 30 | CGT CCT TCC GGC GCA ACG CCT TAT CAT GAT GCT TAT GAA AAT GTG CGT | 624  |
|    | Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asn Val Arg |      |
|    | 195 200 205                                                     |      |
| 35 | GAA GTT ATC CAG CTA CAA GAT CCT GGA CTT GAG CAA CTC AAT GCA TCA | 672  |
|    | Glu Val Ile Gln Leu Gln Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser |      |
|    | 210 215 220                                                     |      |
| 40 | CCG GCA ATT GCC GGG TTG ATG CAT CAA GCC TCC CTA TTG GGT ATT AAC | 720  |
|    | Pro Ala Ile Ala Gly Leu Met His Gln Ala Ser Leu Leu Gly Ile Asn |      |
|    | 225 230 235 240                                                 |      |
|    | GCT TCA ATC TCG CCT GAG CTA TTT AAT ATT CTG ACG GAG GAG ATT ACC | 768  |
|    | Ala Ser Ile Ser Pro Glu Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr |      |
|    | 245 250 255                                                     |      |
| 45 | GAA GGT AAT GCT GAG GAA CTT TAT AAG AAA AAT TTT GGT AAT ATC GAA | 816  |
|    | Glu Gly Asn Ala Glu Glu Leu Tyr Lys Lys Asn Phe Gly Asn Ile Glu |      |
|    | 260 265 270                                                     |      |
| 50 | CCG GCC TCA TTG GCT ATG CCG GAA TAC CTT AAA CGT TAT TAT AAT TTA | 864  |
|    | Pro Ala Ser Leu Ala Met Pro Glu Tyr Leu Lys Arg Tyr Tyr Asn Leu |      |
|    | 275 280 285                                                     |      |
| 55 | AGC GAT GAA GAA CTT AGT CAG TTT ATT GGT AAA GCC AGC AAT TTT GGT | 912  |
|    | Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly Lys Ala Ser Asn Phe Gly |      |
|    | 290 295 300                                                     |      |
| 60 | CAA CAG GAA TAT AGT AAT AAC CAA CTT ATT ACT CCG GTA GTC AAC AGC | 960  |
|    | Gln Gln Glu Tyr Ser Asn Asn Gln Leu Ile Thr Pro Val Val Asn Ser |      |
|    | 305 310 315 320                                                 |      |
|    | AGT GAT GGC ACG GTT AAG GTA TAT CGG ATC ACC CGC GAA TAT ACA ACC | 1008 |
|    | Ser Asp Gly Thr Val Lys Val Tyr Arg Ile Thr Arg Glu Tyr Thr Thr |      |
|    | 325 330 335                                                     |      |
| 65 | AAT GCT TAT CAA ATG GAT GTG GAG CTA TTT CCC TTC GGT GGT GAG AAT | 1056 |
|    | Asn Ala Tyr Gln Met Asp Val Glu Leu Phe Pro Phe Gly Gly Glu Asn |      |
|    | 340 345 350                                                     |      |
| 70 | TAT CCG TTA GAT TAT AAA TTC AAA AAT TTT TAT AAT GCC TCT TAT TTA | 1104 |
|    | Tyr Arg Leu Asp Tyr Lys Phe Lys Asn Phe Tyr Asn Ala Ser Tyr Leu |      |
|    | 355 360 365                                                     |      |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 5  | TCC | ATC | AAG | TTA | AAT | GAT | AAA | AGA | GAA | CTT | GTT | CGA | ACT | GAA | GGC | GCT | 1152 |
|    | Ser | Ile | Lys | Leu | Asn | Asp | Lys | Arg | Glu | Leu | Val | Arg | Thr | Glu | Gly | Ala |      |
|    | 370 |     |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |      |
| 10 | CCT | CAA | GTC | AAT | ATA | GAA | TAC | TCC | GCA | AAT | ATC | ACA | TTA | AAT | ACC | GCT | 1200 |
|    | Pro | Gln | Val | Asn | Ile | Glu | Tyr | Ser | Ala | Asn | Ile | Thr | Leu | Asn | Thr | Ala |      |
|    | 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |      |
| 15 | GAT | ATC | AGT | CAA | CCT | TTT | GAA | ATT | GGC | CTG | ACA | CGA | GTA | CTT | CCT | TCC | 1248 |
|    | Asp | Ile | Ser | Gln | Pro | Phe | Glu | Ile | Gly | Leu | Thr | Arg | Val | Leu | Pro | Ser |      |
|    |     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |      |
| 20 | GGT | TCT | TGG | GCA | TAT | GCC | GCC | GCA | AAA | TTT | ACC | GTT | GAA | GAG | TAT | AAC | 1296 |
|    | Gly | Ser | Trp | Ala | Tyr | Ala | Ala | Ala | Lys | Phe | Thr | Val | Glu | Glu | Tyr | Asn |      |
|    |     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |      |
| 25 | CAA | TAC | TCT | TTT | CTG | CTA | AAA | CTT | AAC | AAG | GCT | ATT | CGT | CTA | TCA | CGT | 1344 |
|    | Gln | Tyr | Ser | Phe | Leu | Leu | Lys | Leu | Asn | Lys | Ala | Ile | Arg | Leu | Ser | Arg |      |
|    |     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |      |
| 30 | GCG | ACA | GAA | TTG | TCA | CCC | ACG | ATT | CTG | GAA | GGC | ATT | GTG | CGC | AGT | GTT | 1392 |
|    | Ala | Thr | Glu | Leu | Ser | Pro | Thr | Ile | Leu | Glu | Gly | Ile | Val | Arg | Ser | Val |      |
|    |     |     | 450 |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |      |
| 35 | AAT | CTA | CAA | CTG | GAT | ATC | AAC | ACA | GAC | GTA | TTA | GGT | AAA | GTT | TTT | CTG | 1440 |
|    | Asn | Leu | Gln | Leu | Asp | Ile | Asn | Thr | Asp | Val | Leu | Gly | Lys | Val | Phe | Leu |      |
|    | 465 |     |     |     |     | 470 |     |     |     | 475 |     |     |     |     | 480 |     |      |
| 40 | ACT | AAA | TAT | TAT | ATG | CAG | CGT | TAT | GCT | ATT | CAT | GCT | GAA | ACT | GCC | CTG | 1488 |
|    | Thr | Lys | Tyr | Tyr | Met | Gln | Arg | Tyr | Ala | Ile | His | Ala | Glu | Thr | Ala | Leu |      |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |      |
| 45 | ATA | CTA | TGC | AAC | GCG | CCT | ATT | TCA | CAA | CGT | TCA | TAT | GAT | AAT | CAA | CCT | 1536 |
|    | Ile | Leu | Cys | Asn | Ala | Pro | Ile | Ser | Gln | Arg | Ser | Tyr | Asp | Asn | Gln | Pro |      |
|    |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |      |
| 50 | AGC | CAA | TTT | GAT | CGC | CTG | TTT | AAT | ACG | CCA | TTA | CTG | AAC | GGA | CAA | TAT | 1584 |
|    | Ser | Gln | Phe | Asp | Arg | Leu | Phe | Asn | Thr | Pro | Leu | Leu | Asn | Gly | Gln | Tyr |      |
|    |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |      |
| 55 | TTT | TCT | ACC | GGC | GAT | GAG | GAG | ATT | GAT | TTA | AAT | TCA | GGT | AGC | ACC | GGC | 1632 |
|    | Phe | Ser | Thr | Gly | Asp | Glu | Glu | Ile | Asp | Leu | Asn | Ser | Gly | Ser | Thr | Gly |      |
|    |     |     | 530 |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |      |
| 60 | GAT | TGG | CGA | AAA | ACC | ATA | CTT | AAG | CGT | GCA | TTT | AAT | ATT | GAT | GAT | GTC | 1680 |
|    | Asp | Trp | Arg | Lys | Thr | Ile | Leu | Lys | Arg | Ala | Phe | Asn | Ile | Asp | Asp | Val |      |
|    |     |     |     |     |     | 550 |     |     |     | 555 |     |     |     |     |     | 560 |      |
| 65 | TCG | CTC | TTC | CGC | CTG | CTT | AAA | ATT | ACC | GAC | CAT | GAT | AAT | AAA | GAT | GGA | 1728 |
|    | Ser | Leu | Phe | Arg | Leu | Leu | Lys | Ile | Thr | Asp | His | Asp | Asn | Lys | Asp | Gly |      |
|    |     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |      |
| 70 | AAA | ATT | AAA | AAT | AAC | CTA | AAG | AAT | CTT | TCC | AAT | TTA | TAT | ATT | GGA | AAA | 1776 |
|    | Lys | Ile | Lys | Asn | Asn | Leu | Lys | Asn | Leu | Ser | Asn | Leu | Tyr | Ile | Gly | Lys |      |
|    |     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |      |
| 75 | TTA | CTG | GCA | GAT | ATT | CAT | CAA | TTA | ACC | ATT | GAT | GAA | CTG | GAT | TTA | TTA | 1824 |
|    | Leu | Leu | Ala | Asp | Ile | His | Gln | Leu | Thr | Ile | Asp | Glu | Leu | Asp | Leu | Leu |      |
|    |     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |      |
| 80 | CTG | ATT | GCC | GTA | GGT | GAA | GGA | AAA | ACT | AAT | TTA | TCC | GCT | ATC | AGT | GAT | 1872 |
|    | Leu | Ile | Ala | Val | Gly | Glu | Gly | Lys | Thr | Asn | Leu | Ser | Ala | Ile | Ser | Asp |      |
|    |     |     | 610 |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |      |
| 85 | AAG | CAA | TTG | GCT | ACC | CTG | ATC | AGA | AAA | CTC | AAT | ACT | ATT | ACC | AGC | TGG | 1920 |
|    | Lys | Gln | Leu | Ala | Thr | Leu | Ile | Arg | Lys | Leu | Asn | Thr | Ile | Thr | Ser | Trp |      |
|    |     |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |      |
| 90 | CTA | CAT | ACA | CAG | AAG | TGG | AGT | GTA | TTC | CAG | CTA | TTT | ATC | ATG | ACC | TCC | 1968 |
|    | Leu | His | Thr | Gln | Lys | Trp | Ser | Val | Phe | Gln | Leu | Phe | Ile | Met | Thr | Ser |      |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    |     |     |     | 545 |     |     |     | 650 |     |     |     | 655 |     |     |     |     |      |
| 5  | ACC | AGC | TAT | AAC | AAA | ACG | CTA | ACG | CCT | GAA | ATT | AAG | AAT | TTG | CTG | GAT | 2015 |
|    | Thr | Ser | Tyr | Asn | Lys | Thr | Leu | Thr | Pro | Glu | Ile | Lys | Asn | Leu | Leu | Asp |      |
|    |     |     |     | 660 |     |     |     | 665 |     |     |     |     |     | 670 |     |     |      |
| 10 | ACC | GTC | TAC | CAC | GGT | TTA | CAA | GGT | TTT | GAT | AAA | GAC | AAA | GCA | GAT | TTG | 2064 |
|    | Thr | Val | Tyr | His | Gly | Leu | Gln | Gly | Phe | Asp | Lys | Asp | Lys | Ala | Asp | Leu |      |
|    |     |     |     | 675 |     |     |     | 680 |     |     |     |     | 685 |     |     |     |      |
| 15 | CTA | CAT | GTC | ATG | GCG | CCC | TAT | ATT | GCG | GCC | ACC | TTG | CAA | TTA | TCA | TCG | 2112 |
|    | Leu | His | Val | Met | Ala | Pro | Tyr | Ile | Ala | Ala | Thr | Leu | Gln | Leu | Ser | Ser |      |
|    |     |     |     | 690 |     |     | 695 |     |     |     |     | 700 |     |     |     |     |      |
| 20 | GAA | AAT | GTC | GCC | CAC | TCG | GTA | CTC | CTT | TGG | GCA | GAT | AAG | TTA | CAG | CCC | 2160 |
|    | Glu | Asn | Val | Ala | His | Ser | Val | Leu | Leu | Trp | Ala | Asp | Lys | Leu | Gln | Pro |      |
|    | 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     | 720 |      |
| 25 | GGC | GAC | GGC | GCA | ATG | ACA | GCA | GAA | AAA | TTC | TGG | GAC | TGG | TTG | AAT | ACT | 2203 |
|    | Gly | Asp | Gly | Ala | Met | Thr | Ala | Glu | Lys | Phe | Trp | Asp | Trp | Leu | Asn | Thr |      |
|    |     |     |     |     | 725 |     |     |     |     | 730 |     |     |     |     | 735 |     |      |
| 30 | AAG | TAT | ACG | CCG | GGT | TCA | TCG | GAA | GCC | GTA | GAA | ACG | CAG | GAA | CAT | ATC | 2256 |
|    | Lys | Tyr | Thr | Pro | Gly | Ser | Ser | Glu | Ala | Val | Glu | Thr | Gln | Glu | His | Ile |      |
|    |     |     |     | 740 |     |     |     | 745 |     |     |     |     |     | 750 |     |     |      |
| 35 | GTT | CAG | TAT | TGT | CAG | GCT | CTG | GCA | CAA | TTG | GAA | ATG | GTT | TAC | CAT | TCC | 2304 |
|    | Val | Gln | Tyr | Cys | Gln | Ala | Leu | Ala | Gln | Leu | Glu | Met | Val | Tyr | His | Ser |      |
|    |     |     |     | 755 |     |     | 760 |     |     |     |     |     | 765 |     |     |     |      |
| 40 | ACC | GGC | ATC | AAC | GAA | AAC | GCC | TTC | CGT | CTA | TTT | GTG | ACA | AAA | CCA | GAG | 2352 |
|    | Thr | Gly | Ile | Asn | Glu | Asn | Ala | Phe | Arg | Leu | Phe | Val | Thr | Lys | Pro | Glu |      |
|    |     | 770 |     |     |     | 775 |     |     |     |     |     | 780 |     |     |     |     |      |
| 45 | ATG | TTT | GGC | GCT | GCA | ACT | GGA | GCA | GCG | CCC | GCG | CAT | GAT | GCC | CTT | TCA | 2400 |
|    | Met | Phe | Gly | Ala | Ala | Thr | Gly | Ala | Ala | Pro | Ala | His | Asp | Ala | Leu | Ser |      |
|    | 785 |     |     |     |     | 790 |     |     |     |     | 795 |     |     |     |     | 800 |      |
| 50 | CTG | ATT | ATG | CTG | ACA | CGT | TTT | GCG | GAT | TGG | GTG | AAC | GCA | CTA | GGC | GAA | 2448 |
|    | Leu | Ile | Met | Leu | Thr | Arg | Phe | Ala | Asp | Trp | Val | Asn | Ala | Leu | Gly | Glu |      |
|    |     |     |     |     | 805 |     |     |     |     | 810 |     |     |     |     | 815 |     |      |
| 55 | AAA | GCG | TCC | TCG | GTG | CTA | GCG | GCA | TTT | GAA | GCT | AAC | TCG | TTA | ACG | GCA | 2496 |
|    | Lys | Ala | Ser | Ser | Val | Leu | Ala | Ala | Phe | Glu | Ala | Asn | Ser | Leu | Thr | Ala |      |
|    |     |     |     | 820 |     |     |     |     | 825 |     |     |     |     | 830 |     |     |      |
| 60 | GAA | CAA | CTG | GCT | GAT | GCC | ATG | AAT | CTT | GAT | GCT | AAT | TTG | CTG | TTG | CAA | 2544 |
|    | Glu | Gln | Leu | Ala | Asp | Ala | Met | Asn | Leu | Asp | Ala | Asn | Leu | Leu | Leu | Gln |      |
|    |     |     | 835 |     |     |     | 840 |     |     |     |     |     | 845 |     |     |     |      |
| 65 | GCC | AGT | ATT | CAA | GCA | CAA | AAT | CAT | CAA | CAT | CTT | CCC | CCA | GTA | ACT | CCA | 2592 |
|    | Ala | Ser | Ile | Gln | Ala | Gln | Asn | His | Gln | His | Leu | Pro | Pro | Val | Thr | Pro |      |
|    |     | 850 |     |     |     |     | 855 |     |     |     |     | 860 |     |     |     |     |      |
| 70 | GAA | AAT | GCG | TTC | TCC | TGT | TGG | ACA | TCT | ATC | AAT | ACT | ATC | CTG | CAA | TGG | 2640 |
|    | Glu | Asn | Ala | Phe | Ser | Cys | Trp | Thr | Ser | Ile | Asn | Thr | Ile | Leu | Gln | Trp |      |
|    | 865 |     |     |     |     | 870 |     |     |     |     | 875 |     |     |     |     | 880 |      |
| 75 | GTT | AAT | GTC | GCA | CAA | CAA | TTG | AAT | GTC | GCC | CCA | CAG | GGC | GTT | TCC | GCT | 2688 |
|    | Val | Asn | Val | Ala | Gln | Gln | Leu | Asn | Val | Ala | Pro | Gln | Gly | Val | Ser | Ala |      |
|    |     |     |     | 885 |     |     |     |     |     | 890 |     |     |     |     | 895 |     |      |
| 80 | TTG | GTC | GGG | CTG | GAT | TAT | ATT | CAA | TCA | ATG | AAA | GAG | ACA | CCG | ACC | TAT | 2736 |
|    | Leu | Val | Gly | Leu | Asp | Tyr | Ile | Gln | Ser | Met | Lys | Glu | Thr | Pro | Thr | Tyr |      |
|    |     |     |     | 900 |     |     |     |     | 905 |     |     |     |     | 910 |     |     |      |
| 85 | GCC | CAG | TGG | GAA | AAC | GCG | GCA | GGC | GTA | TTA | ACC | GCC | GGG | TTG | AAT | TCA | 2784 |
|    | Ala | Gln | Trp | Glu | Asn | Ala | Ala | Gly | Val | Leu | Thr | Ala | Gly | Leu | Asn | Ser |      |
|    |     |     | 915 |     |     |     | 920 |     |     |     |     |     | 925 |     |     |     |      |
| 90 | CAA | CAG | GCT | AAT | ACA | TTA | CAC | GCT | TTT | CTG | GAT | GAA | TCT | CGC | AGT | GCC | 2832 |

|    |     |      |      |      |      |      |      |      |     |      |      |      |      |      |      |      |      |
|----|-----|------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|
|    | Gln | Gln  | Ala  | Asn  | Thr  | Leu  | His  | Ala  | Phe | Leu  | Asp  | Glu  | Ser  | Arg  | Ser  | Ala  |      |
|    | 930 |      |      |      |      |      | 935  |      |     |      |      | 940  |      |      |      |      |      |
| 5  | GCA | TTA  | AGC  | ACC  | TAC  | TAT  | ATC  | CGT  | CAA | GTC  | GCC  | AAG  | GCA  | GCG  | GCG  | GCT  | 2880 |
|    | Ala | Leu  | Ser  | Thr  | Tyr  | Tyr  | Ile  | Arg  | Gln | Val  | Ala  | Lys  | Ala  | Ala  | Ala  | Ala  |      |
|    | 945 |      |      |      |      | 950  |      |      |     |      | 955  |      |      |      |      | 960  |      |
| 10 | ATT | AAA  | AGC  | CGT  | GAT  | GAC  | TTG  | TAT  | CAA | TAC  | TTA  | CTG  | ATT  | GAT  | AAT  | CAG  | 2928 |
|    | Ile | Lys  | Ser  | Arg  | Asp  | Asp  | Leu  | Tyr  | Gln | Tyr  | Leu  | Leu  | Ile  | Asp  | Asn  | Gln  |      |
|    |     |      |      |      | 965  |      |      |      |     | 970  |      |      |      |      | 975  |      |      |
| 15 | GTT | TCT  | GCG  | GCA  | ATA  | AAA  | ACC  | ACC  | CGG | ATC  | GCC  | GAA  | GCC  | ATT  | GCC  | AGT  | 2976 |
|    | Val | Ser  | Ala  | Ala  | Ile  | Lys  | Thr  | Thr  | Arg | Ile  | Ala  | Glu  | Ala  | Ile  | Ala  | Ser  |      |
|    |     |      |      | 980  |      |      |      |      | 985 |      |      |      |      | 990  |      |      |      |
|    | ATT | CAA  | CTG  | TAC  | GTC  | AAC  | CGG  | GCA  | TTG | GAA  | AAT  | GTG  | GAA  | GAA  | AAT  | GCC  | 3024 |
|    | Ile | Gln  | Leu  | Tyr  | Val  | Asn  | Arg  | Ala  | Leu | Glu  | Asn  | Val  | Glu  | Glu  | Asn  | Ala  |      |
|    |     |      | 995  |      |      |      | 1000 |      |     |      |      |      | 1005 |      |      |      |      |
| 20 | AAT | TCG  | GGG  | GTT  | ATC  | AGC  | CGC  | CAA  | TTC | TTT  | ATC  | GAC  | TGG  | GAC  | AAA  | TAC  | 3072 |
|    | Asn | Ser  | Gly  | Val  | Ile  | Ser  | Arg  | Gln  | Phe | Phe  | Ile  | Asp  | Trp  | Asp  | Lys  | Tyr  |      |
|    |     | 1010 |      |      |      |      | 1015 |      |     |      |      | 1020 |      |      |      |      |      |
| 25 | AAT | AAA  | CGC  | TAC  | AGC  | ACT  | TGG  | GCG  | GGT | GTT  | TCT  | CAA  | TTA  | GTT  | TAC  | TAC  | 3120 |
|    | Asn | Lys  | Arg  | Tyr  | Ser  | Thr  | Trp  | Ala  | Gly | Val  | Ser  | Gln  | Leu  | Val  | Tyr  | Tyr  |      |
|    |     | 1025 |      |      |      | 1030 |      |      |     |      | 1035 |      |      |      |      | 1040 |      |
| 30 | CCG | GAA  | AAC  | TAT  | ATT  | GAT  | CCG  | ACC  | ATG | CGT  | ATC  | GGA  | CAA  | ACC  | AAA  | ATG  | 3168 |
|    | Pro | Glu  | Asn  | Tyr  | Ile  | Asp  | Pro  | Thr  | Met | Arg  | Ile  | Gly  | Gln  | Thr  | Lys  | Met  |      |
|    |     |      |      |      | 1045 |      |      |      |     | 1050 |      |      |      |      | 1055 |      |      |
| 35 | ATG | GAC  | GCA  | TTA  | CTG  | CAA  | TCC  | GTC  | AGC | CAA  | AGC  | CAA  | TTA  | AAC  | GCC  | GAT  | 3216 |
|    | Met | Asp  | Ala  | Leu  | Leu  | Gln  | Ser  | Val  | Ser | Gln  | Ser  | Gln  | Leu  | Asn  | Ala  | Asp  |      |
|    |     |      |      | 1060 |      |      |      | 1065 |     |      |      |      |      | 1070 |      |      |      |
|    | ACC | GTC  | GAA  | GAT  | GCC  | TTT  | ATG  | TCT  | TAT | CTG  | ACA  | TCG  | TTT  | GAA  | CAA  | GTG  | 3264 |
|    | Thr | Val  | Glu  | Asp  | Ala  | Phe  | Met  | Ser  | Tyr | Leu  | Thr  | Ser  | Phe  | Glu  | Gln  | Val  |      |
|    |     |      | 1075 |      |      |      | 1080 |      |     |      |      |      | 1085 |      |      |      |      |
| 40 | GCT | AAT  | CTT  | AAA  | GTT  | ATT  | AGC  | GCA  | TAT | CAC  | GAT  | AAT  | ATT  | AAT  | AAC  | GAT  | 3312 |
|    | Ala | Asn  | Leu  | Lys  | Val  | Ile  | Ser  | Ala  | Tyr | His  | Asp  | Asn  | Ile  | Asn  | Asn  | Asp  |      |
|    |     | 1090 |      |      |      | 1095 |      |      |     |      |      | 1100 |      |      |      |      |      |
| 45 | CAA | GGG  | CTG  | ACC  | TAT  | TTT  | ATC  | GGA  | CTC | AGT  | GAA  | ACT  | GAT  | GCC  | GGT  | GAA  | 3360 |
|    | Gln | Gly  | Leu  | Thr  | Tyr  | Phe  | Ile  | Gly  | Leu | Ser  | Glu  | Thr  | Asp  | Ala  | Gly  | Glu  |      |
|    |     | 1105 |      |      |      | 1110 |      |      |     |      | 1115 |      |      |      |      | 1120 |      |
| 50 | TAT | TAT  | TGG  | CGC  | AGT  | GTC  | GAT  | CAC  | AGT | AAA  | TTC  | AAC  | GAC  | GGT  | AAA  | TTC  | 3408 |
|    | Tyr | Tyr  | Trp  | Arg  | Ser  | Val  | Asp  | His  | Ser | Lys  | Phe  | Asn  | Asp  | Gly  | Lys  | Phe  |      |
|    |     |      |      | 1125 |      |      |      |      |     | 1130 |      |      |      | 1135 |      |      |      |
| 55 | GCG | GCT  | AAT  | GCC  | TGG  | AGT  | GAA  | TGG  | CAT | AAA  | ATT  | GAT  | TGT  | CCA  | ATT  | AAC  | 3456 |
|    | Ala | Ala  | Asn  | Ala  | Trp  | Ser  | Glu  | Trp  | His | Lys  | Ile  | Asp  | Cys  | Pro  | Ile  | Asn  |      |
|    |     |      |      | 1140 |      |      |      | 1145 |     |      |      |      | 1150 |      |      |      |      |
|    | CCT | TAT  | AAA  | AGC  | ACT  | ATC  | CGT  | CCA  | GTG | ATA  | TAT  | AAA  | TCC  | CGC  | CTG  | TAT  | 3504 |
|    | Pro | Tyr  | Lys  | Ser  | Thr  | Ile  | Arg  | Pro  | Val | Ile  | Tyr  | Lys  | Ser  | Arg  | Leu  | Tyr  |      |
|    |     |      | 1155 |      |      |      | 1160 |      |     |      |      |      | 1165 |      |      |      |      |
| 60 | CTG | CTC  | TGG  | TTG  | GAA  | CAA  | AAG  | GAG  | ATC | ACC  | AAA  | CAG  | ACA  | GGA  | AAT  | AGT  | 3552 |
|    | Leu | Leu  | Trp  | Leu  | Glu  | Gln  | Lys  | Glu  | Ile | Thr  | Lys  | Gln  | Thr  | Gly  | Asn  | Ser  |      |
|    |     | 1170 |      |      |      | 1175 |      |      |     |      |      | 1180 |      |      |      |      |      |
| 65 | AAA | GAT  | GGC  | TAT  | CAA  | ACT  | GAA  | ACG  | GAT | TAT  | CGT  | TAT  | GAA  | CTA  | AAA  | TTG  | 3600 |
|    | Lys | Asp  | Gly  | Tyr  | Gln  | Thr  | Glu  | Thr  | Asp | Tyr  | Arg  | Tyr  | Glu  | Leu  | Lys  | Leu  |      |
|    |     | 1185 |      |      |      | 1190 |      |      |     |      | 1195 |      |      |      |      | 1200 |      |
| 70 | GCG | CAT  | ATC  | CGC  | TAT  | GAT  | GGC  | ACT  | TGG | AAT  | ACG  | CCA  | ATC  | ACC  | TTT  | GAT  | 3648 |
|    | Ala | His  | Ile  | Arg  | Tyr  | Asp  | Gly  | Thr  | Trp | Asn  | Thr  | Pro  | Ile  | Thr  | Phe  | Asp  |      |
|    |     |      |      | 1205 |      |      |      |      |     | 1210 |      |      |      |      | 1215 |      |      |

|    |      |      |      |      |     |      |      |      |      |      |      |      |      |      |      |      |      |
|----|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|------|------|------|
|    | GTC  | AAT  | AAA  | AAA  | ATA | TCC  | GAG  | CTA  | AAA  | CTG  | GAA  | AAA  | AAT  | AGA  | GCG  | CCC  | 3636 |
|    | Val  | Asn  | Lys  | Lys  | Ile | Ser  | Glu  | Leu  | Lys  | Leu  | Glu  | Lys  | Asn  | Arg  | Ala  | Pro  |      |
|    |      |      |      | 1220 |     |      |      |      | 1225 |      |      |      |      | 1230 |      |      |      |
| 5  | GGA  | CTC  | TAT  | TGT  | GCC | GGT  | TAT  | CAA  | GGT  | GAA  | GAT  | ACG  | TTG  | CTG  | GTG  | ATG  | 3744 |
|    | Gly  | Leu  | Tyr  | Cys  | Ala | Gly  | Tyr  | Gln  | Gly  | Glu  | Asp  | Thr  | Leu  | Leu  | Val  | Met  |      |
|    |      |      | 1235 |      |     |      |      | 1240 |      |      |      |      | 1245 |      |      |      |      |
| 10 | TTT  | TAT  | AAC  | CAA  | CAA | GAC  | ACA  | CTA  | GAT  | AGT  | TAT  | AAA  | AAC  | GCT  | TCA  | ATG  | 3792 |
|    | Phe  | Tyr  | Asn  | Gln  | Gln | Asp  | Thr  | Leu  | Asp  | Ser  | Tyr  | Lys  | Asn  | Ala  | Ser  | Met  |      |
|    |      | 1250 |      |      |     |      | 1255 |      |      |      |      | 1260 |      |      |      |      |      |
| 15 | CAA  | GGA  | CTA  | TAT  | ATC | TTT  | GCT  | GAT  | ATG  | GCA  | TCC  | AAA  | GAT  | ATG  | ACC  | CCA  | 3840 |
|    | Gln  | Gly  | Leu  | Tyr  | Ile | Phe  | Ala  | Asp  | Met  | Ala  | Ser  | Lys  | Asp  | Met  | Thr  | Pro  |      |
|    | 1265 |      |      |      |     | 1270 |      |      |      |      | 1275 |      |      |      |      | 1280 |      |
| 20 | GAA  | CAG  | AGC  | AAT  | GTT | TAT  | CGG  | GAT  | AAT  | AGC  | TAT  | CAA  | CAA  | TTT  | GAT  | ACC  | 3888 |
|    | Glu  | Gln  | Ser  | Asn  | Val | Tyr  | Arg  | Asp  | Asn  | Ser  | Tyr  | Gln  | Gln  | Phe  | Asp  | Thr  |      |
|    |      |      |      | 1285 |     |      |      |      |      | 1290 |      |      |      |      | 1295 |      |      |
| 25 | AAT  | AAT  | GTC  | AGA  | AGA | GTG  | AAT  | AAC  | CGC  | TAT  | GCA  | GAG  | GAT  | TAT  | GAG  | ATT  | 3936 |
|    | Asn  | Asn  | Val  | Arg  | Arg | Val  | Asn  | Asn  | Arg  | Tyr  | Ala  | Glu  | Asp  | Tyr  | Glu  | Ile  |      |
|    |      |      |      | 1300 |     |      |      |      | 1305 |      |      |      |      |      | 1310 |      |      |
| 30 | CCT  | TCC  | TCG  | GTA  | AGT | AGC  | CGT  | AAA  | GAC  | TAT  | GGT  | TGG  | GGA  | GAT  | TAT  | TAC  | 3984 |
|    | Pro  | Ser  | Ser  | Val  | Ser | Ser  | Arg  | Lys  | Asp  | Tyr  | Gly  | Trp  | Gly  | Asp  | Tyr  | Tyr  |      |
|    |      |      | 1315 |      |     |      |      | 1320 |      |      |      |      | 1325 |      |      |      |      |
| 35 | CTC  | AGC  | ATG  | GTA  | TAT | AAC  | GGA  | GAT  | ATT  | CCA  | ACT  | ATC  | AAT  | TAC  | AAA  | GCC  | 4032 |
|    | Leu  | Ser  | Met  | Val  | Tyr | Asn  | Gly  | Asp  | Ile  | Pro  | Thr  | Ile  | Asn  | Tyr  | Lys  | Ala  |      |
|    |      | 1330 |      |      |     |      | 1335 |      |      |      |      | 1340 |      |      |      |      |      |
| 40 | GCA  | TCA  | AGT  | GAT  | TTA | AAA  | ATC  | TAT  | ATC  | TCA  | CCA  | AAA  | TTA  | AGA  | ATT  | ATT  | 4080 |
|    | Ala  | Ser  | Ser  | Asp  | Leu | Lys  | Ile  | Tyr  | Ile  | Ser  | Pro  | Lys  | Leu  | Arg  | Ile  | Ile  |      |
|    | 1345 |      |      |      |     | 1350 |      |      |      | 1355 |      |      |      |      |      | 1360 |      |
| 45 | CAT  | AAT  | GGA  | TAT  | GAA | GGA  | CAG  | AAG  | CGC  | AAT  | CAA  | TGC  | AAT  | CTG  | ATG  | AAT  | 4128 |
|    | His  | Asn  | Gly  | Tyr  | Glu | Gly  | Gln  | Lys  | Arg  | Asn  | Gln  | Cys  | Asn  | Leu  | Met  | Asn  |      |
|    |      |      |      | 1365 |     |      |      |      |      | 1370 |      |      |      |      | 1375 |      |      |
| 50 | AAA  | TAT  | GGC  | AAA  | CTA | GGT  | GAT  | AAA  | TTT  | ATT  | GTT  | TAT  | ACT  | AGC  | TTG  | GGG  | 4176 |
|    | Lys  | Tyr  | Gly  | Lys  | Leu | Gly  | Asp  | Lys  | Phe  | Ile  | Val  | Tyr  | Thr  | Ser  | Leu  | Gly  |      |
|    |      |      | 1380 |      |     |      |      |      | 1385 |      |      |      |      | 1390 |      |      |      |
| 55 | GTC  | AAT  | CCA  | AAT  | AAC | TCG  | TCA  | AAT  | AAG  | CTC  | ATG  | TTT  | TAC  | CCC  | GTC  | TAT  | 4224 |
|    | Val  | Asn  | Pro  | Asn  | Asn | Ser  | Ser  | Asn  | Lys  | Leu  | Met  | Phe  | Tyr  | Pro  | Val  | Tyr  |      |
|    |      |      | 1395 |      |     |      |      | 1400 |      |      |      |      | 1405 |      |      |      |      |
| 60 | CAA  | TAT  | AGC  | GGA  | AAC | ACC  | AGT  | GGA  | CTC  | AAT  | CAA  | GGG  | AGA  | CTA  | CTA  | TTC  | 4272 |
|    | Gln  | Tyr  | Ser  | Gly  | Asn | Thr  | Ser  | Gly  | Leu  | Asn  | Gln  | Gly  | Arg  | Leu  | Leu  | Phe  |      |
|    |      | 1410 |      |      |     |      | 1415 |      |      |      |      | 1420 |      |      |      |      |      |
| 65 | CAC  | CGT  | GAC  | ACC  | ACT | TAT  | CCA  | TCT  | AAA  | GTA  | GAA  | GCT  | TGG  | ATT  | CCT  | GGA  | 4320 |
|    | His  | Arg  | Asp  | Thr  | Thr | Tyr  | Pro  | Ser  | Lys  | Val  | Glu  | Ala  | Trp  | Ile  | Pro  | Gly  |      |
|    | 1425 |      |      |      |     | 1430 |      |      |      |      | 1435 |      |      |      |      | 1440 |      |
| 70 | GCA  | AAA  | CGT  | TCT  | CTA | ACC  | AAC  | CAA  | AAT  | GCC  | GCC  | ATT  | GGT  | GAT  | GAT  | TAT  | 4368 |
|    | Ala  | Lys  | Arg  | Ser  | Leu | Thr  | Asn  | Gln  | Asn  | Ala  | Ala  | Ile  | Gly  | Asp  | Asp  | Tyr  |      |
|    |      |      |      | 1445 |     |      |      |      |      | 1450 |      |      |      |      | 1455 |      |      |
| 75 | GCT  | ACA  | GAC  | TCT  | CTG | AAT  | AAA  | CCG  | GAT  | GAT  | CTT  | AAG  | CAA  | TAT  | ATC  | TTT  | 4416 |
|    | Ala  | Thr  | Asp  | Ser  | Leu | Asn  | Lys  | Pro  | Asp  | Asp  | Leu  | Lys  | Gln  | Tyr  | Ile  | Phe  |      |
|    |      |      | 1460 |      |     |      |      | 1465 |      |      |      |      |      | 1470 |      |      |      |
| 80 | ATG  | ACT  | GAC  | AGT  | AAA | GGG  | ACT  | GCT  | ACT  | GAT  | GTC  | TCA  | GGC  | CCA  | GTA  | GAG  | 4464 |
|    | Met  | Thr  | Asp  | Ser  | Lys | Gly  | Thr  | Ala  | Thr  | Asp  | Val  | Ser  | Gly  | Pro  | Val  | Glu  |      |
|    |      |      | 1475 |      |     |      | 1480 |      |      |      |      | 1485 |      |      |      |      |      |
| 85 | ATT  | AAT  | ACT  | GCA  | ATT | TCT  | CCA  | GCA  | AAA  | GTT  | CAG  | ATA  | ATA  | GTC  | AAA  | GCG  | 4512 |
|    | Ile  | Asn  | Thr  | Ala  | Ile | Ser  | Pro  | Ala  | Lys  | Val  | Gln  | Ile  | Ile  | Val  | Lys  | Ala  |      |
|    |      | 1490 |      |      |     |      | 1495 |      |      |      |      | 1500 |      |      |      |      |      |

|    |                                                                                     |
|----|-------------------------------------------------------------------------------------|
| 5  | GGT GGC AAG GAG CAA ACT TTT ACC GCA GAT AAA GAT GTC TCC ATT CAG 4560                |
|    | Gly Gly Lys Glu Gln Thr Phe Thr Ala Asp Lys Asp Val Ser Ile Gln 1505 1510 1515 1520 |
| 10 | CCA TCA CCT AGC TTT GAT GAA ATG AAT TAT CAA TTT AAT GCC CTT GAA 4608                |
|    | Pro Ser Pro Ser Phe Asp Glu Met Asn Tyr Gln Phe Asn Ala Leu Glu 1525 1530 1535      |
| 15 | ATA GAC GGT TCT GGT CTG AAT TTT ATT AAC AAC TCA GCC AGT ATT GAT 4656                |
|    | Ile Asp Gly Ser Gly Leu Asn Phe Ile Asn Asn Ser Ala Ser Ile Asp 1540 1545 1550      |
| 20 | GTT ACT TTT ACC GCA TTT GCG GAG GAT GGC CGC AAA CTG GGT TAT GAA 4704                |
|    | Val Thr Phe Thr Ala Phe Ala Glu Asp Gly Arg Lys Leu Gly Tyr Glu 1555 1560 1565      |
| 25 | AGT TTC AGT ATT CCT GTT ACC CTC AAG GTA AGT ACC GAT AAT GCC CTG 4752                |
|    | Ser Phe Ser Ile Pro Val Thr Leu Lys Val Ser Thr Asp Asn Ala Leu 1570 1575 1580      |
| 30 | ACC CTG CAC CAT AAT GAA AAT GGT GCG CAA TAT ATG CAA TGG CAA TCC 4800                |
|    | Thr Leu His His Asn Glu Asn Gly Ala Gln Tyr Met Gln Trp Gln Ser 1585 1590 1595 1600 |
| 35 | TAT CGT ACC CGC CTG AAT ACT CTA TTT GCC CGC CAG TTG GTT GCA CGC 4848                |
|    | Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Arg Gln Leu Val Ala Arg 1605 1610 1615      |
| 40 | GCC ACC ACC GGA ATC GAT ACA ATT CTG AGT ATG GAA ACT CAG AAT ATT 4896                |
|    | Ala Thr Thr Gly Ile Asp Thr Ile Leu Ser Met Glu Thr Gln Asn Ile 1620 1625 1630      |
| 45 | CAG GAA CCG CAG TTA GGC AAA GGT TTC TAT GCT ACG TTC GTG ATA CCT 4944                |
|    | Gln Glu Pro Gln Leu Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro 1635 1640 1645      |
| 50 | CCC TAT AAC CTA TCA ACT CAT GGT GAT GAA CGT TGG TTT AAG CTT TAT 4992                |
|    | Pro Tyr Asn Leu Ser Thr His Gly Asp Glu Arg Trp Phe Lys Leu Tyr 1650 1655 1660      |
| 55 | ATC AAA CAT GTT GTT GAT AAT AAT TCA CAT ATT ATC TAT TCA GGC CAG 5040                |
|    | Ile Lys His Val Val Asp Asn Asn Ser His Ile Ile Tyr Ser Gly Gln 1665 1670 1675 1680 |
| 60 | CTA ACA GAT ACA AAT ATA AAC ATC ACA TTA TTT ATT CCT CTT GAT GAT 5088                |
|    | Leu Thr Asp Thr Asn Ile Asn Ile Thr Leu Phe Ile Pro Leu Asp Asp 1685 1690 1695      |
| 65 | GTC CCA TTG AAT CAA GAT TAT CAC GCC AAG GTT TAT ATG ACC TTC AAG 5136                |
|    | Val Pro Leu Asn Gln Asp Tyr His Ala Lys Val Tyr Met Thr Phe Lys 1700 1705 1710      |
| 70 | AAA TCA CCA TCA GAT GGT ACC TGG TGG GGC CCT CAC TTT GTT AGA GAT 5184                |
|    | Lys Ser Pro Ser Asp Gly Thr Trp Trp Gly Pro His Phe Val Arg Asp 1715 1720 1725      |
| 75 | GAT AAA GGA ATA GTA ACA ATA AAC CCT AAA TCC ATT TTG ACC CAT TTT 5232                |
|    | Asp Lys Gly Ile Val Thr Ile Asn Pro Lys Ser Ile Leu Thr His Phe 1730 1735 1740      |
| 80 | GAG AGC GTC AAT GTC CTG AAT AAT ATT AGT AGC GAA CCA ATG GAT TTC 5280                |
|    | Glu Ser Val Asn Val Leu Asn Asn Ile Ser Ser Glu Pro Met Asp Phe 1745 1750 1755 1760 |
| 85 | AGC GGC GCT AAC AGC CTC TAT TTC TGG GAA CTG TTC TAC TAT ACC CCG 5328                |
|    | Ser Gly Ala Asn Ser Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro 1765 1770 1775      |
| 90 | ATG CTG GTT GCT CAA CGT TTG CTG CAT GAA CAG AAC TTC GAT GAA GCC 5376                |
|    | Met Leu Val Ala Gln Arg Leu Leu His Glu Gln Asn Phe Asp Glu Ala                     |

|    | 1730                                                                                                                                   | 1785                                                                               | 1790                               |      |
|----|----------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|------------------------------------|------|
| 5  | AAC CGT TGG CTG AAA TAT GTC<br>Asn Arg Trp Leu Lys Tyr Val                                                                             | TGG AGT CCA TCC GGT<br>Trp S r Pro Ser Gly                                         | TAT ATT GTC CAC<br>Tyr Ile Val His | 5424 |
|    | 1795                                                                                                                                   | 1800                                                                               | 1805                               |      |
| 10 | GGC CAG ATT CAG AAC TAC<br>Gly Gln Ile Gln Asn Tyr                                                                                     | CAG TGG AAC GTC CGC CCG TTA CTG GAA GAC<br>Gln Trp Asn Val Arg Pro Leu Leu Glu Asp |                                    | 5472 |
|    | 1810                                                                                                                                   | 1815                                                                               | 1820                               |      |
| 15 | ACC AGT TGG AAC AGT GAT CCT TTG GAT TCC GTC GAT CCT GAC GCG GTA<br>Thr Ser Trp Asn Ser Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val     |                                                                                    |                                    | 5520 |
|    | 1825                                                                                                                                   | 1830                                                                               | 1835                               | 1840 |
| 20 | GCA CAG CAC GAT CCA ATG CAC TAC AAA GTT TCA ACT TTT ATG CGT ACC<br>Ala Gln His Asp Pro Met His Tyr Lys Val Ser Thr Phe Met Arg Thr     |                                                                                    |                                    | 5568 |
|    | 1845                                                                                                                                   | 1850                                                                               | 1855                               |      |
| 25 | TTG GAT CTA TTG ATA GCA CGC GGC GAC CAT GCT TAT CGC CAA CTG GAA<br>Leu Asp Leu Leu Ile Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu     |                                                                                    |                                    | 5616 |
|    | 1860                                                                                                                                   | 1865                                                                               | 1870                               |      |
| 30 | CGA GAT ACA CTC AAC GAA GCG AAG ATG TGG TAT ATG CAA GCG CTG CAT<br>Arg Asp Thr Leu Asn Glu Ala Lys Met Trp Tyr Met Gln Ala Leu His     |                                                                                    |                                    | 5664 |
|    | 1875                                                                                                                                   | 1880                                                                               | 1885                               |      |
| 35 | CTA TTA GGT GAC AAA CCT TAT CTA CCG CTG AGT ACG ACA TGG AGT GAT<br>Leu Leu Gly Asp Lys Pro Tyr Leu Pro Leu Ser Thr Thr Trp Ser Asp     |                                                                                    |                                    | 5712 |
|    | 1890                                                                                                                                   | 1895                                                                               | 1900                               |      |
| 40 | CCA CGA CTA GAC AGA GCC GCG GAT ATC ACT ACC CAA AAT GCT CAC GAC<br>Pro Arg Leu Asp Arg Ala Ala Asp Ile Thr Thr Thr Gln Asn Ala His Asp |                                                                                    |                                    | 5760 |
|    | 1905                                                                                                                                   | 1910                                                                               | 1915                               | 1920 |
| 45 | AGC GCA ATA GTC GCT CTG CGG CAG AAT ATA CCT ACA CCG GCA CCT TTA<br>Ser Ala Ile Val Ala Leu Arg Gln Asn Ile Pro Thr Pro Ala Pro Leu     |                                                                                    |                                    | 5808 |
|    | 1925                                                                                                                                   | 1930                                                                               | 1935                               |      |
| 50 | TCA TTG CGC AGC GCT AAT ACC CTG ACT GAT CTC TTC CTG CCG CAA ATC<br>Ser Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile     |                                                                                    |                                    | 5856 |
|    | 1940                                                                                                                                   | 1945                                                                               | 1950                               |      |
| 55 | AAT GAA GTG ATG ATG AAT TAC TGG CAG ACA TTA GCT CAG AGA GTA TAC<br>Asn Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr     |                                                                                    |                                    | 5904 |
|    | 1955                                                                                                                                   | 1960                                                                               | 1965                               |      |
| 60 | AAT CTG CGT CAT AAC CTC TCT ATC GAC GGC CAG CCG TTA TAT CTG CCA<br>Asn Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro     |                                                                                    |                                    | 5952 |
|    | 1970                                                                                                                                   | 1975                                                                               | 1980                               |      |
| 65 | ATC TAT GCC ACA CCG GCC GAT CCG AAA GCG TTA CTC AGC GCC GCC GTT<br>Ile Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val     |                                                                                    |                                    | 6000 |
|    | 1985                                                                                                                                   | 1990                                                                               | 1995                               | 2000 |
| 70 | GCC ACT TCT CAA GGT GGA GGC AAG CTA CCG GAA TCA TTT ATG TCC CTG<br>Ala Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu     |                                                                                    |                                    | 6048 |
|    | 2005                                                                                                                                   | 2010                                                                               | 2015                               |      |
| 75 | TGG CGT TTC CCG CAC ATG CTG GAA AAT GCG CGC GGC ATG GTT AGC CAG<br>Trp Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln     |                                                                                    |                                    | 6096 |
|    | 2020                                                                                                                                   | 2025                                                                               | 2030                               |      |
| 80 | CTC ACC CAG TTC GGC TCC ACG TTA CAA AAT ATT ATC GAA CGT CAG GAC<br>Leu Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp     |                                                                                    |                                    | 6144 |
|    | 2035                                                                                                                                   | 2040                                                                               | 2045                               |      |
| 85 | GCG GAA GCG CTC AAT GCG TTA TTA CAA AAT CAG GCC GCC GAG CTG ATA<br>Ala Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile     |                                                                                    |                                    | 6192 |
|    | 2050                                                                                                                                   | 2055                                                                               | 2060                               |      |
| 90 | TTG ACT AAC CTG AGC ATT CAG GAC AAA ACC ATT GAA GAA TTG GAT GCC<br>Leu Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu     |                                                                                    |                                    | 6240 |
|    | 2065                                                                                                                                   | 2070                                                                               | 2075                               |      |

|    |      |     |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|----|------|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
|    | Leu  | Thr | Asn | Leu  | Ser  | Ile  | Gln  | Asp  | Lys  | Thr  | Ile  | Glu  | Glu  | Leu  | Asp  | Ala  |      |
|    | 2055 |     |     |      |      | 2070 |      |      |      |      | 2075 |      |      |      |      | 2080 |      |
| 5  | GAG  | AAA | ACG | GTG  | TTG  | GAA  | AAA  | TCC  | AAA  | GCG  | GGA  | GCA  | CAA  | TCG  | CGC  | TTT  | 6283 |
|    | Glu  | Lys | Thr | Val  | Leu  | Glu  | Lys  | Ser  | Lys  | Ala  | Gly  | Ala  | Gln  | Ser  | Arg  | Phe  |      |
|    |      |     |     |      | 2085 |      |      |      |      | 2090 |      |      |      |      | 2095 |      |      |
| 10 | GAT  | AGC | TAC | GGC  | AAA  | CTG  | TAC  | GAT  | GAG  | AAT  | ATC  | AAC  | GCC  | GGT  | GAA  | AAC  | 6336 |
|    | Asp  | Ser | Tyr | Gly  | Lys  | Leu  | Tyr  | Asp  | Glu  | Asn  | Ile  | Asn  | Ala  | Gly  | Glu  | Asn  |      |
|    |      |     |     | 2100 |      |      |      |      | 2105 |      |      |      |      | 2110 |      |      |      |
| 15 | CAA  | GCC | ATG | ACG  | CTA  | CGA  | GCG  | TCC  | GCC  | GCC  | GGG  | CTT  | ACC  | ACG  | GCA  | GTT  | 6384 |
|    | Gln  | Ala | Met | Thr  | Leu  | Arg  | Ala  | Ser  | Ala  | Ala  | Gly  | Leu  | Thr  | Thr  | Ala  | Val  |      |
|    |      |     |     | 2115 |      |      |      | 2120 |      |      |      |      | 2125 |      |      |      |      |
| 20 | CAG  | GCA | TCC | CGT  | CTG  | GCC  | GGT  | GCG  | GCG  | GCT  | GAT  | CTG  | GTG  | CCT  | AAC  | ATC  | 6432 |
|    | Gln  | Ala | Ser | Arg  | Leu  | Ala  | Ala  | Gly  | Ala  | Ala  | Asp  | Leu  | Val  | Pro  | Asn  | Ile  |      |
|    |      |     |     | 2130 |      |      | 2135 |      |      |      |      | 2140 |      |      |      |      |      |
| 25 | TTC  | GGC | TTT | GCC  | GGT  | GGC  | GGC  | AGC  | CGT  | TGG  | GGG  | GCT  | ATC  | GCT  | GAG  | GCG  | 6480 |
|    | Phe  | Gly | Phe | Ala  | Gly  | Gly  | Gly  | Ser  | Arg  | Trp  |      | Gly  | Ala  | Ile  | Ala  | Glu  | Ala  |
|    |      |     |     |      |      | 2150 |      |      |      |      | 2155 |      |      |      |      | 2160 |      |
| 30 | ACA  | GGT | TAT | GTG  | ATG  | GAA  | TTC  | TCC  | GCG  | AAT  | GTT  | ATG  | AAC  | ACC  | GAA  | GCG  | 6528 |
|    | Thr  | Gly | Tyr | Val  | Met  | Glu  | Phe  | Ser  | Ala  | Asn  | Val  | Met  | Asn  | Thr  | Glu  | Ala  |      |
|    |      |     |     |      | 2165 |      |      |      |      | 2170 |      |      |      |      | 2175 |      |      |
| 35 | GAT  | AAA | ATT | AGC  | CAA  | TCT  | GAA  | ACC  | TAC  | CGT  | CGT  | CGC  | CGT  | CAG  | GAG  | TGG  | 6576 |
|    | Asp  | Lys | Ile | Ser  | Gln  | Ser  | Glu  | Thr  | Tyr  | Arg  | Arg  | Arg  | Arg  | Gln  | Glu  | Trp  |      |
|    |      |     |     | 2180 |      |      |      |      | 2185 |      |      |      |      | 2190 |      |      |      |
| 40 | GAG  | ATC | CAG | CGG  | AAT  | AAT  | GCC  | GAA  | GCG  | GAA  | TTG  | AAG  | CAA  | ATC  | GAT  | GCT  | 6624 |
|    | Glu  | Ile | Gln | Arg  | Asn  | Asn  | Ala  | Glu  | Ala  | Glu  | Leu  | Lys  | Gln  | Ile  | Asp  | Ala  |      |
|    |      |     |     | 2195 |      |      |      | 2200 |      |      |      |      | 2205 |      |      |      |      |
| 45 | CAG  | CTC | AAA | TCA  | CTC  | GCT  | GTA  | CGC  | CGC  | GAA  | GCC  | GCC  | GTA  | TTG  | CAG  | AAA  | 6672 |
|    | Gln  | Leu | Lys | Ser  | Leu  | Ala  | Val  | Arg  | Arg  | Glu  | Ala  | Ala  | Val  | Leu  | Gln  | Lys  |      |
|    |      |     |     | 2210 |      |      | 2215 |      |      |      |      | 2220 |      |      |      |      |      |
| 50 | ACC  | AGT | CTG | AAA  | ACC  | CAA  | CAA  | GAA  | CAG  | ACC  | CAA  | TCT  | CAA  | TTG  | GCC  | TTC  | 6720 |
|    | Thr  | Ser | Leu | Lys  | Thr  | Gln  | Gln  | Glu  | Gln  | Thr  | Gln  | Ser  | Gln  | Leu  | Ala  | Phe  |      |
|    |      |     |     |      |      | 2230 |      |      |      |      | 2235 |      |      |      | 2240 |      |      |
| 55 | CTG  | CAA | CGT | AAG  | TTC  | AGC  | AAT  | CAG  | GCG  | TTA  | TAC  | AAC  | TGG  | CTG  | CGT  | GGT  | 6768 |
|    | Leu  | Gln | Arg | Lys  | Phe  | Ser  | Asn  | Gln  | Ala  | Leu  | Tyr  | Asn  | Trp  | Leu  | Arg  | Gly  |      |
|    |      |     |     |      | 2245 |      |      |      |      | 2250 |      |      |      |      | 2255 |      |      |
| 60 | CGA  | CTG | GCG | GCG  | ATT  | TAC  | TTC  | CAG  | TTC  | TAC  | GAT  | TTG  | GCC  | GTC  | GCG  | CGT  | 6816 |
|    | Arg  | Leu | Ala | Ala  | Ile  | Tyr  | Phe  | Gln  | Phe  | Tyr  | Asp  | Leu  | Ala  | Val  | Ala  | Arg  |      |
|    |      |     |     | 2260 |      |      |      |      | 2265 |      |      |      |      | 2270 |      |      |      |
| 65 | TGC  | CTG | ATG | GCA  | GAA  | CAA  | GCT  | TAC  | CGT  | TGG  | GAA  | CTC  | AAT  | GAT  | GAC  | TCT  | 6864 |
|    | Cys  | Leu | Met | Ala  | Glu  | Gln  | Ala  | Tyr  | Arg  | Trp  | Glu  | Leu  | Asn  | Asp  | Asp  | Ser  |      |
|    |      |     |     | 2275 |      |      |      | 2280 |      |      |      |      | 2285 |      |      |      |      |
| 70 | GCC  | GCG | TTC | ATT  | AAA  | CCG  | GGC  | GCC  | TGG  | CAG  | GGA  | ACC  | TAT  | GCC  | GGT  | CTG  | 6912 |
|    | Ala  | Arg | Phe | Ile  | Lys  | Pro  | Gly  | Ala  | Trp  | Gln  | Gly  | Thr  | Tyr  | Ala  | Gly  | Leu  |      |
|    |      |     |     | 2290 |      |      | 2295 |      |      |      |      | 2300 |      |      |      |      |      |
| 75 | CTT  | GCA | GGT | GAA  | ACC  | TTG  | ATG  | CTG  | AGT  | CTG  | GCA  | CAA  | ATG  | GAA  | GAC  | GCT  | 6960 |
|    | Leu  | Ala | Gly | Glu  | Thr  | Leu  | Met  | Leu  | Ser  | Leu  | Ala  | Gln  | Met  | Glu  | Asp  | Ala  |      |
|    |      |     |     |      |      | 2310 |      |      |      |      | 2315 |      |      |      | 2320 |      |      |
| 80 | CAT  | CTG | AAA | GCG  | GAT  | AAA  | GCG  | GCA  | TTA  | GAG  | GTT  | GAA  | GCG  | ACA  | GTA  | TCG  | 7008 |
|    | His  | Leu | Lys | Arg  | Asp  | Lys  | Arg  | Ala  | Leu  | Glu  | Val  | Glu  | Arg  | Thr  | Val  | Ser  |      |
|    |      |     |     |      | 2325 |      |      |      |      | 2330 |      |      |      |      | 2335 |      |      |
| 85 | CTG  | GCC | GAA | GTT  | TAT  | GCA  | GGA  | TTA  | CCA  | AAA  | GAT  | AAC  | GGT  | CCA  | TTT  | TCC  | 7056 |
|    | Leu  | Ala | Glu | Val  | Tyr  | Ala  | Gly  | Leu  | Pro  | Lys  | Asp  | Asn  | Gly  | Pro  | Phe  | Ser  |      |
|    |      |     |     | 2340 |      |      |      |      | 2345 |      |      |      |      | 2350 |      |      |      |



CTG GCT CAG GAA ATT GAC AAG CTG GTG AGT CAA GST TCA GGC AGT GCC 7174  
 Leu Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala  
 2355 2360 2365

5 GGC AGT GGT AAT AAT AAT TTG GCG TTC GGC GCC GGC ACG GAC ACT AAA 7152  
 Gly Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys  
 2370 2375 2380

10 ACC TCT TTG CAG GCA TCA GTT TCA TTC GCT GAT TTG AAA ATT CGT GAA 7200  
 Thr Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu  
 2385 2390 2395 2400

15 GAT TAC CCG GCA TCG CTT GGC AAA ATT CGA CGT ATC AAA CAG ATC AGC 7248  
 Asp Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser  
 2405 2410 2415

20 GTC ACT TTG CCC GCG CTA CTG GGA CCG TAT CAG GAT GTA CAG GCA ATA 7296  
 Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile  
 2420 2425 2430

25 TTG TCT TAC GGC GAT AAA GCC GGA TTA GCT AAC GGC TGT GAA GCG CTG 7344  
 Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu  
 2435 2440 2445

30 GCA GTT TCT CAC GGT ATG AAT GAC AGC GGC CAA TTC CAG CTC GAT TTC 7392  
 Ala Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe  
 2450 2455 2460

35 AAC GAT GGC AAA TTC CTG CCA TTC GAA GGC ATC GCC ATT GAT CAA GGC 7440  
 Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly  
 2465 2470 2475 2480

40 ACG CTG ACA CTG AGC TTC CCA AAT GCA TCT ATG CCG GAG AAA GGT AAA 7488  
 Thr Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys  
 2485 2490 2495

45 CAA GCC ACT ATG TTA AAA ACC CTG AAC GAT ATC ATT TTG CAT ATT CGC 7536  
 Gln Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg  
 2500 2505 2510

TAC ACC ATT AAA TAA 7551  
 Tyr Thr Ile Lys ...  
 2516

## (2) INFORMATION FOR SEQ ID NO:47:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2516 amino acids  
 (B) TYPE: amino acids  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47 (TcdA):

| Features | From | To   | Description                  |
|----------|------|------|------------------------------|
| Peptide  | 1    | 2516 | TcdA proteins                |
| Peptide  | 89   | 1937 | TcdA <sub>111</sub> peptide  |
| Fragment | 89   | 100  | S2 N-terminus (SEQ ID NO:13) |
| Fragment | 284  | 299  | (SEQ ID NO:38)               |
| Fragment | 554  | 563  | (SEQ ID NO:17)               |
| Fragment | 1080 | 1092 | (SEQ ID NO:23; 12/13)        |
| Fragment | 1385 | 1400 | (SEQ ID NO:18)               |
| Fragment | 1478 | 1497 | (SEQ ID NO:39)               |
| Fragment | 1620 | 1642 | (SEQ ID NO:21; 19/23)        |
| Fragment | 1938 | 1948 | (SEQ ID NO:41)               |
| Peptide  | 1938 | 2516 | TcdA <sub>111</sub> peptide  |
| Fragment | 2327 | 2345 | (SEQ ID NO:42)               |
| Fragment | 2398 | 2408 | (SEQ ID NO:43)               |

Met Asn Glu Ser Val Lys Glu Ile Pro Asp Val Leu Lys Ser Gln Cys  
 1 5 10 15  
 5 Gly Phe Asn Cys Leu Thr Asp Il Ser His Ser Ser Phe Asn Glu Phe  
 20 25 30  
 Arg Gln Gln Val Ser Glu His Leu Ser Trp Ser Glu Thr His Asp Leu  
 35 40 45  
 10 Tyr His Asp Ala Gln Gln Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala  
 50 55 60  
 15 Arg Ile Leu Lys Arg Ala Asn Pro Gln Leu Gln Asn Ala Val His Leu  
 65 70 75 80  
 Ala Ile Leu Ala Pro Asn Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe  
 85 90 95  
 20 Ser Gly Arg Ala Ser Gln Tyr Val Ala Pro Gly Thr Val Ser Ser Met  
 100 105 110  
 Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn  
 115 120 125  
 25 Leu His Ala Ser Asp Ser Val Tyr Tyr Leu Asp Thr Arg Arg Pro Asp  
 130 135 140  
 30 Leu Lys Ser Met Ala Leu Ser Gln Gln Asn Met Asp Ile Glu Leu Ser  
 145 150 155 160  
 Thr Leu Ser Leu Ser Asn Glu Leu Leu Leu Glu Ser Ile Lys Thr Glu  
 165 170 175  
 35 Ser Lys Leu Glu Asn Tyr Thr Lys Val Met Glu Met Leu Ser Thr Phe  
 180 185 190  
 Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asn Val Arg  
 195 200 205  
 40 Glu Val Ile Gln Leu Gln Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser  
 210 215 220  
 45 Pro Ala Ile Ala Gly Leu Met His Gln Ala Ser Leu Leu Gly Ile Asn  
 225 230 235 240  
 Ala Ser Ile Ser Pro Glu Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr  
 245 250 255  
 50 Glu Gly Asn Ala Glu Glu Leu Tyr Lys Lys Asn Phe Gly Asn Ile Glu  
 260 265 270  
 Pro Ala Ser Leu Ala Met Pro Glu Tyr Leu Lys Arg Tyr Tyr Asn Leu  
 275 280 285  
 55 Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly Lys Ala Ser Asn Phe Gly  
 290 295 300  
 60 Gln Gln Glu Tyr Ser Asn Asn Gln Leu Ile Thr Pro Val Val Asn Ser  
 305 310 315 320  
 Ser Asp Gly Thr Val Lys Val Tyr Arg Ile Thr Arg Glu Tyr Thr Thr  
 325 330 335  
 65 Asn Ala Tyr Gln Met Asp Val Glu Leu Phe Pro Phe Gly Gly Glu Asn  
 340 345 350  
 Tyr Arg Leu Asp Tyr Lys Phe Lys Asn Phe Tyr Asn Ala Ser Tyr Leu  
 355 360 365  
 70 Ser Ile Lys Leu Asn Asp Lys Arg Glu Leu Val Arg Thr Glu Gly Ala

|    | 370                                                             | 375     | 380     |
|----|-----------------------------------------------------------------|---------|---------|
|    | Pro Gln Val Asn Ile Glu Tyr Ser Ala Asn Ile Thr Leu Asn Thr Ala |         |         |
|    | 385                                                             | 390     | 395 400 |
| 5  | Asp Ile Ser Gln Pro Phe Glu Ile Gly Leu Thr Arg Val Leu Pro Ser |         |         |
|    |                                                                 | 405 410 | 415     |
| 10 | Gly Ser Trp Ala Tyr Ala Ala Ala Lys Phe Thr Val Glu Glu Tyr Asn |         |         |
|    |                                                                 | 420 425 | 430     |
|    | Gln Tyr Ser Phe Leu Leu Lys Leu Asn Lys Ala Ile Arg Leu Ser Arg |         |         |
|    |                                                                 | 435 440 | 445     |
| 15 | Ala Thr Glu Leu Ser Pro Thr Ile Leu Glu Gly Ile Val Arg Ser Val |         |         |
|    |                                                                 | 450 455 | 460     |
|    | Asn Leu Gln Leu Asp Ile Asn Thr Asp Val Leu Gly Lys Val Phe Leu |         |         |
|    |                                                                 | 465 470 | 475 480 |
| 20 | Thr Lys Tyr Tyr Met Gln Arg Tyr Ala Ile His Ala Glu Thr Ala Leu |         |         |
|    |                                                                 | 485 490 | 495     |
| 25 | Ile Leu Cys Asn Ala Pro Ile Ser Gln Arg Ser Tyr Asp Asn Gln Pro |         |         |
|    |                                                                 | 500 505 | 510     |
|    | Ser Gln Phe Asp Arg Leu Phe Asn Thr Pro Leu Leu Asn Gly Gln Tyr |         |         |
|    |                                                                 | 515 520 | 525     |
| 30 | Phe Ser Thr Gly Asp Glu Glu Ile Asp Leu Asn Ser Gly Ser Thr Gly |         |         |
|    |                                                                 | 530 535 | 540     |
|    | Asp Trp Arg Lys Thr Ile Leu Lys Arg Ala Phe Asn Ile Asp Asp Val |         |         |
|    |                                                                 | 545 550 | 555 560 |
| 35 | Ser Leu Phe Arg Leu Leu Lys Ile Thr Asp His Asp Asn Lys Asp Gly |         |         |
|    |                                                                 | 565 570 | 575     |
| 40 | Lys Ile Lys Asn Asn Leu Lys Asn Leu Ser Asn Leu Tyr Ile Gly Lys |         |         |
|    |                                                                 | 580 585 | 590     |
|    | Leu Leu Ala Asp Ile His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu |         |         |
|    |                                                                 | 595 600 | 605     |
| 45 | Leu Ile Ala Val Gly Glu Glu Lys Thr Asn Leu Ser Ala Ile Ser Asp |         |         |
|    |                                                                 | 610 615 | 620     |
|    | Lys Gln Leu Ala Thr Leu Ile Arg Lys Leu Asn Thr Ile Thr Ser Trp |         |         |
|    |                                                                 | 625 630 | 635 640 |
| 50 | Leu His Thr Gln Lys Trp Ser Val Phe Gln Leu Phe Ile Met Thr Ser |         |         |
|    |                                                                 | 645 650 | 655     |
| 55 | Thr Ser Tyr Asn Lys Thr Leu Thr Pro Glu Ile Lys Asn Leu Leu Asp |         |         |
|    |                                                                 | 660 665 | 670     |
|    | Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asp Leu |         |         |
|    |                                                                 | 675 680 | 685     |
| 60 | Leu His Val Met Ala Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser |         |         |
|    |                                                                 | 690 695 | 700     |
|    | Glu Asn Val Ala His Ser Val Leu Leu Trp Ala Asp Lys Leu Gln Pro |         |         |
|    |                                                                 | 705 710 | 715 720 |
| 65 | Gly Asp Gly Ala Met Thr Ala Glu Lys Phe Trp Asp Trp Leu Asn Thr |         |         |
|    |                                                                 | 725 730 | 735     |
| 70 | Lys Tyr Thr Pro Gly Ser Ser Glu Ala Val Glu Thr Gln Glu His Ile |         |         |
|    |                                                                 | 740 745 | 750     |

Val Gln Tyr Cys Gln Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser  
 755 760 765  
 5 Thr Gly Ile Asn Glu Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu  
 770 775 780  
 Met Phe Gly Ala Ala Thr Gly Ala Ala Pro Ala His Asp Ala Leu Ser  
 785 790 795 800  
 10 Leu Ile Met Leu Thr Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu  
 805 810 815  
 Lys Ala Ser Ser Val Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala  
 820 825 830  
 15 Glu Gln Leu Ala Asp Ala Met Asn Leu Asp Ala Asn Leu Leu Gln  
 835 840 845  
 Ala Ser Ile Gln Ala Gln Asn His Gln His Leu Pro Pro Val Thr Pro  
 850 855 860  
 Glu Asn Ala Phe Ser Cys Trp Thr Ser Ile Asn Thr Ile Leu Gln Trp  
 865 870 875 880  
 25 Val Asn Val Ala Gln Gln Leu Asn Val Ala Pro Gln Gly Val Ser Ala  
 885 890 895  
 Leu Val Gly Leu Asp Tyr Ile Gln Ser Met Lys Glu Thr Pro Thr Tyr  
 900 905 910  
 30 Ala Gln Trp Glu Asn Ala Ala Gly Val Leu Thr Ala Gly Leu Asn Ser  
 915 920 925  
 Gln Gln Ala Asn Thr Leu His Ala Phe Leu Asp Glu Ser Arg Ser Ala  
 930 935 940  
 35 Ala Leu Ser Thr Tyr Tyr Ile Arg Gln Val Ala Lys Ala Ala Ala Ala  
 945 950 955 960  
 Ile Lys Ser Arg Asp Asp Leu Tyr Gln Tyr Leu Leu Ile Asp Asn Gln  
 965 970 975  
 Val Ser Ala Ala Ile Lys Thr Thr Arg Ile Ala Glu Ala Ile Ala Ser  
 980 985 990  
 45 Ile Gln Leu Tyr Val Asn Arg Ala Leu Glu Asn Val Glu Glu Asn Ala  
 995 1000 1005  
 Asn Ser Gly Val Ile Ser Arg Gln Phe Phe Ile Asp Trp Asp Lys Tyr  
 1010 1015 1020  
 50 Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Gln Leu Val Tyr Tyr  
 1025 1030 1035 1040  
 Pro Glu Asn Tyr Ile Asp Pro Thr Met Arg Ile Gly Gln Thr Lys Met  
 1045 1050 1055  
 Met Asp Ala Leu Leu Gln Ser Val Ser Gln Ser Gln Leu Asn Ala Asp  
 1060 1065 1070  
 60 Thr Val Glu Asp Ala Phe Met Ser Tyr Leu Thr Ser Phe Glu Gln Val  
 1075 1080 1085  
 Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Ile Asn Asn Asp  
 1090 1095 1100  
 65 Gln Gly Leu Thr Tyr Phe Ile Gly Leu Ser Glu Thr Asp Ala Gly Glu  
 1105 1110 1115 1120  
 Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Phe Asn Asp Gly Lys Phe  
 1125 1130 1135

Ala Ala Asn Ala Trp Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asn  
1140 1145 1150

5 Pro Tyr Lys Ser Thr Ile Arg Pro Val Ile Tyr Lys Ser Arg Leu Tyr  
1155 1160 1165

Leu Leu Trp Leu Glu Gln Lys Glu Ile Thr Lys Gln Thr Gly Asn Ser  
1170 1175 1180

10 Lys Asp Gly Tyr Gln Thr Glu Thr Asp Tyr Arg Tyr Glu Leu Lys Leu  
1185 1190 1195 1200

Ala His Ile Arg Tyr Asp Gly Thr Trp Asn Thr Pro Ile Thr Phe Asp  
1205 1210 1215

15 Val Asn Lys Lys Ile Ser Glu Leu Lys Leu Glu Lys Asn Arg Ala Pro  
1220 1225 1230

20 Gly Leu Tyr Cys Ala Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met  
1235 1240 1245

Phe Tyr Asn Gln Gln Asp Thr Leu Asp Ser Tyr Lys Asn Ala Ser Met  
1250 1255 1260

25 Gln Gly Leu Tyr Ile Phe Ala Asp Met Ala Ser Lys Asp Met Thr Pro  
1265 1270 1275 1280

Glu Gln Ser Asn Val Tyr Arg Asp Asn Ser Tyr Gln Gln Phe Asp Thr  
1285 1290 1295

30 Asn Asn Val Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr Glu Ile  
1300 1305 1310

35 Pro Ser Ser Val Ser Ser Arg Lys Asp Tyr Gly Trp Gly Asp Tyr Tyr  
1315 1320 1325

Leu Ser Met Val Tyr Asn Gly Asp Ile Pro Thr Ile Asn Tyr Lys Ala  
1330 1335 1340

40 Ala Ser Ser Asp Leu Lys Ile Tyr Ile Ser Pro Lys Leu Arg Ile Ile  
1345 1350 1355 1360

His Asn Gly Tyr Glu Gly Gln Lys Arg Asn Gln Cys Asn Leu Met Asn  
1365 1370 1375

Lys Tyr Gly Lys Leu Gly Asp Lys Phe Ile Val Tyr Thr Ser Leu Gly  
1380 1385 1390

50 Val Asn Pro Asn Asn Ser Ser Asn Lys Leu Met Phe Tyr Pro Val Tyr  
1395 1400 1405

Gln Tyr Ser Gly Asn Thr Ser Gly Leu Asn Gln Gly Arg Leu Leu Phe  
1410 1415 1420

55 His Arg Asp Thr Thr Tyr Pro Ser Lys Val Glu Ala Trp Ile Pro Gly  
1425 1430 1435 1440

Ala Lys Arg Ser Leu Thr Asn Gln Asn Ala Ala Ile Gly Asp Asp Tyr  
1445 1450 1455

Ala Thr Asp Ser Leu Asn Lys Pro Asp Asp Leu Lys Gln Tyr Ile Phe  
1460 1465 1470

65 Met Thr Asp Ser Lys Gly Thr Ala Thr Asp Val Ser Gly Pro Val Glu  
1475 1480 1485

Ile Asn Thr Ala Ile Ser Pro Ala Lys Val Gln Ile Ile Val Lys Ala  
1490 1495 1500

70 Gly Gly Lys Glu Gln Thr Phe Thr Ala Asp Lys Asp Val Ser Ile Gln

|    |                                                                 |      |      |  |      |  |      |
|----|-----------------------------------------------------------------|------|------|--|------|--|------|
|    | 1505                                                            |      | 1510 |  | 1515 |  | 1520 |
|    | Pro Ser Pro Ser Phe Asp Glu Met Asn Tyr Gln Phe Asn Ala Leu Glu |      |      |  |      |  |      |
|    |                                                                 | 1525 |      |  | 1530 |  | 1535 |
| 5  | Ile Asp Gly Ser Gly L u Asn Phe Ile Asn Asn Ser Ala Ser Ile Asp |      |      |  |      |  |      |
|    |                                                                 | 1540 |      |  | 1545 |  | 1550 |
| 10 | Val Thr Phe Thr Ala Phe Ala Glu Asp Gly Arg Lys Leu Gly Tyr Glu |      |      |  |      |  |      |
|    |                                                                 | 1555 |      |  | 1560 |  | 1565 |
|    | Ser Phe Ser Ile Pro Val Thr Leu Lys Val Ser Thr Asp Asn Ala Leu |      |      |  |      |  |      |
|    |                                                                 | 1570 |      |  | 1575 |  | 1580 |
| 15 | Thr Leu His His Asn Glu Asn Gly Ala Gln Tyr Met Gln Trp Gln Ser |      |      |  |      |  |      |
|    |                                                                 | 1585 |      |  | 1590 |  | 1595 |
|    | Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Arg Gln Leu Val Ala Arg |      |      |  |      |  |      |
|    |                                                                 |      | 1605 |  | 1610 |  | 1615 |
| 20 | Ala Thr Thr Gly Ile Asp Thr Ile Leu Ser Met Glu Thr Gln Asn Ile |      |      |  |      |  |      |
|    |                                                                 | 1620 |      |  | 1625 |  | 1630 |
| 25 | Gln Glu Pro Gln Leu Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro |      |      |  |      |  |      |
|    |                                                                 | 1635 |      |  | 1640 |  | 1645 |
|    | Pro Tyr Asn Leu Ser Thr His Gly Asp Glu Arg Trp Phe Lys Leu Tyr |      |      |  |      |  |      |
|    |                                                                 | 1650 |      |  | 1655 |  | 1660 |
| 30 | Ile Lys His Val Val Asp Asn Asn Ser His Ile Ile Tyr Ser Gly Gln |      |      |  |      |  |      |
|    |                                                                 | 1665 |      |  | 1670 |  | 1675 |
|    | Leu Thr Asp Thr Asn Ile Asn Ile Thr Leu Phe Ile Pro Leu Asp Asp |      |      |  |      |  |      |
|    |                                                                 |      | 1685 |  | 1690 |  | 1695 |
| 35 | Val Pro Leu Asn Gln Asp Tyr His Ala Lys Val Tyr Met Thr Phe Lys |      |      |  |      |  |      |
|    |                                                                 | 1700 |      |  | 1705 |  | 1710 |
| 40 | Lys Ser Pro Ser Asp Gly Thr Trp Trp Gly Pro His Phe Val Arg Asp |      |      |  |      |  |      |
|    |                                                                 | 1715 |      |  | 1720 |  | 1725 |
|    | Asp Lys Gly Ile Val Thr Ile Asn Pro Lys Ser Ile Leu Thr His Phe |      |      |  |      |  |      |
|    |                                                                 | 1730 |      |  | 1735 |  | 1740 |
| 45 | Glu Ser Val Asn Val Leu Asn Asn Ile Ser Ser Glu Pro Met Asp Phe |      |      |  |      |  |      |
|    |                                                                 | 1745 |      |  | 1750 |  | 1755 |
|    | Ser Gly Ala Asn Ser Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro |      |      |  |      |  |      |
|    |                                                                 |      | 1765 |  | 1770 |  | 1775 |
| 50 | Met Leu Val Ala Gln Arg Leu Leu His Glu Gln Asn Phe Asp Glu Ala |      |      |  |      |  |      |
|    |                                                                 | 1780 |      |  | 1785 |  | 1790 |
| 55 | Asn Arg Trp Leu Lys Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val His |      |      |  |      |  |      |
|    |                                                                 | 1795 |      |  | 1800 |  | 1805 |
|    | Gly Gln Ile Gln Asn Tyr Gln Trp Asn Val Arg Pro Leu Leu Glu Asp |      |      |  |      |  |      |
|    |                                                                 | 1810 |      |  | 1815 |  | 1820 |
| 60 | Thr Ser Trp Asn Ser Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val |      |      |  |      |  |      |
|    |                                                                 | 1825 |      |  | 1830 |  | 1835 |
|    | Ala Gln His Asp Pro Met His Tyr Lys Val Ser Thr Phe Met Arg Thr |      |      |  |      |  |      |
|    |                                                                 |      | 1845 |  | 1850 |  | 1855 |
| 65 | Leu Asp Leu Leu Ile Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu |      |      |  |      |  |      |
|    |                                                                 | 1860 |      |  | 1865 |  | 1870 |
| 70 | Arg Asp Thr Leu Asn Glu Ala Lys Met Trp Tyr Met Gln Ala Leu His |      |      |  |      |  |      |
|    |                                                                 | 1875 |      |  | 1880 |  | 1885 |

Leu Leu Gly Asp Lys Pro Tyr Leu Pro Leu Ser Thr Thr Trp Ser Asp  
 1890 1895 1900  
 5 Pro Arg Leu Asp Arg Ala Ala Asp Ile Thr Thr Gln Asn Ala His Asp  
 1905 1910 1915 1920  
 Ser Ala Ile Val Ala Leu Arg Gln Asn Ile Pro Thr Pro Ala Pro Leu  
 1925 1930 1935  
 10 Ser Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile  
 1940 1945 1950  
 Asn Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr  
 1955 1960 1965  
 15 Asn Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro  
 1970 1975 1980  
 Ile Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val  
 1985 1990 1995 2000  
 Ala Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu  
 2005 2010 2015  
 25 Trp Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln  
 2020 2025 2030  
 Leu Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp  
 2035 2040 2045  
 30 Ala Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile  
 2050 2055 2060  
 Leu Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala  
 2065 2070 2075 2080  
 Glu Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe  
 2085 2090 2095  
 40 Asp Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Gly Glu Asn  
 2100 2105 2110  
 Gln Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val  
 2115 2120 2125  
 45 Gln Ala Ser Arg Leu Ala Gly Ala Ala Ala Asp Leu Val Pro Asn Ile  
 2130 2135 2140  
 Phe Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala  
 2145 2150 2155 2160  
 Thr Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala  
 2165 2170 2175  
 55 Asp Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp  
 2180 2185 2190  
 Glu Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala  
 2195 2200 2205  
 60 Gln Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys  
 2210 2215 2220  
 Thr Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe  
 2225 2230 2235 2240  
 Leu Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly  
 2245 2250 2255  
 70 Arg Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg  
 2260 2265 2270

Cys Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser  
 2275 2280 2285  
 5 Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu  
 2290 2295 2300  
 Leu Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala  
 2305 2310 2315 2320  
 10 His Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser  
 2325 2330 2335  
 15 Leu Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser  
 2340 2345 2350  
 Leu Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala  
 2355 2360 2365  
 20 Gly Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys  
 2370 2375 2380  
 Thr Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu  
 2385 2390 2395 2400  
 25 Asp Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser  
 2405 2410 2415  
 30 Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile  
 2420 2425 2430  
 Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu  
 2435 2440 2445  
 35 Ala Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe  
 2450 2455 2460  
 Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly  
 2465 2470 2475 2480  
 40 Thr Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys  
 2485 2490 2495  
 45 Gln Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg  
 2500 2505 2510  
 Tyr Thr Ile Lys  
 2516

50

## (2) INFORMATION FOR SEQ ID NO:48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5547 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

## (ii) MOLECULE TYPE: DNA (genomic)

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48 (tcdA<sub>ii</sub> coding region):

CTG ATA GGC TAT AAC AAT CAA TTT AGC GGT AGA GCC AGT CAA TAT GTT 48  
 Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val  
 1 5 10 15  
 GCG CCG GGT ACC GTT TCT TCC ATG TTC TCC CCC GCC GCT TAT TTG ACT 96  
 Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr L u Thr  
 20 25 30

65



|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | GAA | CTT | TAT | CGT | GAA | GCA | CGC | AAT | TTA | CAC | GCA | AGT | GAC | TCC | GTT | TAT | 144 |
|    | Glu | Leu | Tyr | Arg | Glu | Ala | Arg | Asn | Leu | His | Ala | Ser | Asp | Ser | Val | Tyr |     |
| 5  |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |
|    | TAT | CTG | GAT | ACC | CGC | CGC | CCA | GAT | CTC | AAA | TCA | ATG | GCG | CTC | AGT | CAG | 192 |
|    | Tyr | Leu | Asp | Thr | Arg | Arg | Pro | Asp | Leu | Lys | Ser | Met | Ala | Leu | Ser | Gln |     |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |     |
| 10 | CAA | AAT | ATG | GAT | ATA | GAA | TTA | TCC | ACA | CTC | TCT | TTG | TCC | AAT | GAG | CTG | 240 |
|    | Gln | Asn | Met | Asp | Ile | Glu | Leu | Ser | Thr | Leu | Ser | Leu | Ser | Asn | Glu | Leu |     |
|    |     | 65  |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |     |
| 15 | TTA | TTG | GAA | AGC | ATT | AAA | ACT | GAA | TCT | AAA | CTG | GAA | AAC | TAT | ACT | AAA | 288 |
|    | Leu | Leu | Glu | Ser | Ile | Lys | Thr | Glu | Ser | Lys | Leu | Glu | Asn | Tyr | Thr | Lys |     |
|    |     |     |     | 85  |     |     |     |     |     | 90  |     |     |     |     | 95  |     |     |
| 20 | GTG | ATG | GAA | ATG | CTC | TCC | ACT | TTC | CGT | CCT | TCC | GGC | GCA | ACG | CCT | TAT | 336 |
|    | Val | Met | Glu | Met | Leu | Ser | Thr | Phe | Arg | Pro | Ser | Gly | Ala | Thr | Pro | Tyr |     |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |     |
| 25 | CAT | GAT | GCT | TAT | GAA | AAT | GTG | CGT | GAA | GTT | ATC | CAG | CTA | CAA | GAT | CCT | 384 |
|    | His | Asp | Ala | Tyr | Glu | Asn | Val | Arg | Glu | Val | Ile | Gln | Leu | Gln | Asp | Pro |     |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
|    | GGA | CTT | GAG | CAA | CTC | AAT | GCA | TCA | CCG | GCA | ATT | GCC | GGG | TTG | ATG | CAT | 432 |
|    | Gly | Leu | Glu | Gln | Leu | Asn | Ala | Ser | Pro | Ala | Ile | Ala | Gly | Leu | Met | His |     |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |     |
| 30 | CAA | GCC | TCC | CTA | TTG | GGT | ATT | AAC | GCT | TCA | ATC | TCG | CCT | GAG | CTA | TTT | 480 |
|    | Gln | Ala | Ser | Leu | Leu | Gly | Ile | Asn | Ala | Ser | Ile | Ser | Pro | Glu | Leu | Phe |     |
|    |     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |     |
| 35 | AAT | ATT | CTG | ACG | GAG | GAG | ATT | ACC | GAA | GGT | AAT | GCT | GAG | GAA | CTT | TAT | 528 |
|    | Asn | Ile | Leu | Thr | Glu | Glu | Ile | Thr | Glu | Gly | Asn | Ala | Glu | Glu | Leu | Tyr |     |
|    |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |     |
| 40 | AAG | AAA | AAT | TTT | GGT | AAT | ATC | GAA | CCG | GCC | TCA | TTG | GCT | ATG | CCG | GAA | 576 |
|    | Lys | Lys | Asn | Phe | Gly | Asn | Ile | Glu | Pro | Ala | Ser | Leu | Ala | Met | Pro | Glu |     |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |
| 45 | TAC | CTT | AAA | CGT | TAT | TAT | AAT | TTA | AGC | GAT | GAA | GAA | CTT | AGT | CAG | TTT | 624 |
|    | Tyr | Leu | Lys | Arg | Tyr | Tyr | Asn | Leu | Ser | Asp | Glu | Glu | Leu | Ser | Gln | Phe |     |
|    |     |     | 195 |     |     |     | 200 |     |     |     |     |     | 205 |     |     |     |     |
|    | ATT | GGT | AAA | GCC | AGC | AAT | TTT | GGT | CAA | CAG | GAA | TAT | AGT | AAT | AAC | CAA | 672 |
|    | Ile | Gly | Lys | Ala | Ser | Asn | Phe | Gly | Gln | Gln | Glu | Tyr | Ser | Asn | Asn | Gln |     |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |     |
| 50 | CTT | ATT | ACT | CCG | GTA | GTC | AAC | AGC | AGT | GAT | GGC | ACG | GTT | AAG | GTA | TAT | 720 |
|    | Leu | Ile | Thr | Pro | Val | Val | Asn | Ser | Ser | Asp | Gly | Thr | Val | Lys | Val | Tyr |     |
|    |     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |     |
| 55 | CGG | ATC | ACC | CGC | GAA | TAT | ACA | ACC | AAT | GCT | TAT | CAA | ATG | GAT | GTG | GAG | 768 |
|    | Arg | Ile | Thr | Arg | Glu | Tyr | Thr | Thr | Asn | Ala | Tyr | Gln | Met | Asp | Val | Glu |     |
|    |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |     |
| 60 | CTA | TTT | CCC | TTC | GGT | GGT | GAG | AAT | TAT | CGG | TTA | GAT | TAT | AAA | TTC | AAA | 816 |
|    | Leu | Phe | Pro | Phe | Gly | Gly | Glu | Asn | Tyr | Arg | Leu | Asp | Tyr | Lys | Phe | Lys |     |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |     |
| 65 | AAT | TTT | TAT | AAT | GCC | TCT | TAT | TTA | TCC | ATC | AAG | TTA | AAT | GAT | AAA | AGA | 864 |
|    | Asn | Phe | Tyr | Asn | Ala | Ser | Tyr | Leu | Ser | Ile | Lys | Leu | Asn | Asp | Lys | Arg |     |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |     |
|    | GAA | CTT | GTT | CGA | ACT | GAA | GGC | GCT | CCT | CAA | GTC | AAT | ATA | GAA | TAC | TCC | 912 |
|    | Glu | Leu | Val | Arg | Thr | Glu | Gly | Ala | Pro | Gln | Val | Asn | Ile | Glu | Tyr | Ser |     |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |     |
| 70 | GCA | AAT | ATC | ACA | TTA | AAT | ACC | GCT | GAT | ATC | AGT | CAA | CCT | TTT | GAA | ATT | 960 |
|    | Ala | Asn | Ile | Thr | Leu | Asn | Thr | Ala | Asp | Ile | Ser | Gln | Pro | Phe | Glu | Ile |     |

|    |                                                                      |     |     |     |     |  |     |     |
|----|----------------------------------------------------------------------|-----|-----|-----|-----|--|-----|-----|
|    | 305                                                                  |     | 310 |     | 315 |  | 320 |     |
|    | GGC CTG ACA CGA GTA CTT CCT TCC GGT TCT TGG GCA TAT GCC GCC GCA 1008 |     |     |     |     |  |     |     |
| 5  | Gly Leu Thr Arg Val Leu Pro Ser Gly Ser Trp Ala Tyr Ala Ala Ala      | 325 | 330 | 335 |     |  |     |     |
|    | AAA TTT ACC GTT GAA GAG TAT AAC CAA TAC TCT TTT CTG CTA AAA CTT 1056 |     |     |     |     |  |     |     |
|    | Lys Phe Thr Val Glu Glu Tyr Asn Gln Tyr Ser Phe Leu Leu Lys Leu      | 340 | 345 | 350 |     |  |     |     |
| 10 | AAC AAG GCT ATT CGT CTA TCA CGT GCG ACA GAA TTG TCA CCC ACC ATT 1104 |     |     |     |     |  |     |     |
|    | Asn Lys Ile Arg Leu Ser Arg Ala Thr Glu Leu Ser Pro Thr Ile          | 355 | 360 | 365 |     |  |     |     |
| 15 | CTG GAA GGC ATT GTG CGC AGT GTT AAT CTA CAA CTG GAT ATC AAC ACA 1152 |     |     |     |     |  |     |     |
|    | Leu Glu Gly Ile Val Arg Ser Val Asn Leu Gln Leu Asp Ile Asn Thr      | 370 | 375 | 380 |     |  |     |     |
| 20 | GAC GTA TTA GGT AAA GTT TTT CTG ACT AAA TAT TAT ATG CAG CGT TAT 1200 |     |     |     |     |  |     |     |
|    | Asp Val Leu Gly Lys Phe Leu Thr Lys Tyr Tyr Met Gln Arg Tyr          | 385 | 390 | 395 |     |  |     | 400 |
| 25 | GCT ATT CAT GCT GAA ACT GCC CTG ATA CTA TGC AAC GCG CCT ATT TCA 1248 |     |     |     |     |  |     |     |
|    | Ala Ile His Ala Glu Thr Ala Leu Ile Leu Cys Asn Ala Pro Ile Ser      | 405 | 410 | 415 |     |  |     |     |
| 30 | CAA CGT TCA TAT GAT AAT CAA CCT AGC CAA TTT GAT CGC CTG TTT AAT 1296 |     |     |     |     |  |     |     |
|    | Gln Arg Ser Tyr Asp Asn Gln Pro Ser Gln Phe Asp Arg Leu Phe Asn      | 420 | 425 | 430 |     |  |     |     |
|    | ACG CCA TTA CTG AAC GGA CAA TAT TTT TCT ACC GGC GAT GAG GAG ATT 1344 |     |     |     |     |  |     |     |
|    | Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp Glu Glu Ile      | 435 | 440 | 445 |     |  |     |     |
| 35 | GAT TTA AAT TCA GGT AGC ACC GGC GAT TGG CGA AAA ACC ATA CTT AAG 1392 |     |     |     |     |  |     |     |
|    | Asp Leu Asn Ser Gly Ser Thr Gly Asp Trp Arg Lys Thr Ile Leu Lys      | 450 | 455 | 460 |     |  |     |     |
| 40 | CGT GCA TTT AAT ATT GAT GAT GTC TCG CTC TTC CGC CTG CTT AAA ATT 1440 |     |     |     |     |  |     |     |
|    | Arg Ala Phe Asn Ile Asp Asp Val Ser Leu Phe Arg Leu Leu Lys Ile      | 465 | 470 | 475 |     |  |     | 480 |
| 45 | ACC GAC CAT GAT AAT AAA GAT GGA AAA ATT AAA AAT AAC CTA AAG AAT 1488 |     |     |     |     |  |     |     |
|    | Thr Asp His Asp Asn Lys Asp Gly Lys Ile Lys Asn Asn Leu Lys Asn      | 485 | 490 | 495 |     |  |     |     |
| 50 | CTT TCC AAT TTA TAT ATT GGA AAA TTA CTG GCA GAT ATT CAT CAA TTA 1536 |     |     |     |     |  |     |     |
|    | Leu Ser Asn Leu Tyr Ile Gly Lys Leu Leu Ala Asp Ile His Gln Leu      | 500 | 505 | 510 |     |  |     |     |
|    | ACC ATT GAT GAA CTG GAT TTA TTA CTG ATT GCC GTA GGT GAA GGA AAA 1584 |     |     |     |     |  |     |     |
|    | Thr Ile Asp Glu Leu Asp Leu Leu Ile Ala Val Gly Glu Gly Lys          | 515 | 520 | 525 |     |  |     |     |
| 55 | ACT AAT TTA TCC GCT ATC AGT GAT AAG CAA TTG GCT ACC CTG ATC AGA 1632 |     |     |     |     |  |     |     |
|    | Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Thr Leu Ile Arg      | 530 | 535 | 540 |     |  |     |     |
| 60 | AAA CTC AAT ACT ATT ACC AGC TGG CTA CAT ACA CAG AAG TGG AGT GTA 1680 |     |     |     |     |  |     |     |
|    | Lys Leu Asn Thr Ile Thr Ser Trp Leu His Thr Gln Lys Trp Ser Val      | 545 | 550 | 555 |     |  |     | 560 |
| 65 | TTC CAG CTA TTT ATC ATG ACC TCC ACC AGC TAT AAC AAA ACG CTA ACG 1728 |     |     |     |     |  |     |     |
|    | Phe Gln Leu Phe Ile Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr      | 565 | 570 | 575 |     |  |     |     |
| 70 | CCT GAA ATT AAG AAT TTG CTG GAT ACC GTC TAC CAC GGT TTA CAA GGT 1776 |     |     |     |     |  |     |     |
|    | Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly      | 580 | 585 | 590 |     |  |     |     |
|    | TTT GAT AAA GAC AAA GCA GAT TTG CTA CAT GTC ATG GCG CCC TAT ATT 1824 |     |     |     |     |  |     |     |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | Phe | Asp | Lys | Asp | Lys | Ala | Asp | Leu | Leu | His | Val | Met | Ala | Pro | Tyr | Ile |      |
|    |     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |      |
| 5  | GCG | GCC | ACC | TTG | CAA | TTA | TCA | TCG | GAA | AAT | GTC | GCC | CAC | TCG | GTA | CTC | 1372 |
|    | Ala | Ala | Thr | Leu | Gln | Leu | Ser | Ser | Glu | Asn | Val | Ala | His | Ser | Val | Leu |      |
|    |     | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |      |
| 10 | CTT | TGG | GCA | GAT | AAG | TTA | CAG | CCC | GGC | GAC | GGC | GCA | ATG | ACA | GCA | GAA | 1920 |
|    | Leu | Trp | Ala | Asp | Lys | Leu | Gln | Pro | Gly | Asp | Gly | Ala | Met | Thr | Ala | Glu |      |
|    | 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |      |
| 15 | AAA | TTC | TGG | GAC | TGG | TTG | AAT | ACT | AAG | TAT | ACG | CCG | GGT | TCA | TCG | GAA | 1963 |
|    | Lys | Phe | Trp | Asp | Trp | Leu | Asn | Thr | Lys | Tyr | Thr | Pro | Gly | Ser | Ser | Glu |      |
|    |     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |      |
| 20 | GCC | GTA | GAA | ACG | CAG | GAA | CAT | ATC | GTT | CAG | TAT | TGT | CAG | GCT | CTG | GCA | 2016 |
|    | Ala | Val | Glu | Thr | Gln | Glu | His | Ile | Val | Gln | Tyr | Cys | Gln | Ala | Leu | Ala |      |
|    |     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |      |
| 25 | CAA | TTG | GAA | ATG | GTT | TAC | CAT | TCC | ACC | GGC | ATC | AAC | GAA | AAC | GCC | TTC | 2064 |
|    | Gln | Leu | Glu | Met | Val | Tyr | His | Ser | Thr | Gly | Ile | Asn | Glu | Asn | Ala | Phe |      |
|    |     |     | 675 |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |      |
| 30 | CGT | CTA | TTT | GTG | ACA | AAA | CCA | GAG | ATG | TTT | GGC | GCT | GCA | ACT | GGA | GCA | 2112 |
|    | Arg | Leu | Phe | Val | Thr | Lys | Pro | Glu | Met | Phe | Gly | Ala | Ala | Thr | Gly | Ala |      |
|    |     | 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |      |
| 35 | GCG | CCC | GCG | CAT | GAT | GCC | CTT | TCA | CTG | ATT | ATG | CTG | ACA | CGT | TTT | GCG | 2160 |
|    | Ala | Pro | Ala | His | Asp | Ala | Leu | Ser | Leu | Ile | Met | Leu | Thr | Arg | Phe | Ala |      |
|    | 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     | 720 |      |
| 40 | GAT | TGG | GTG | AAC | GCA | CTA | GGC | GAA | AAA | GCG | TCC | TCG | GTG | CTA | GCG | GCA | 2208 |
|    | Asp | Trp | Val | Asn | Ala | Leu | Gly | Glu | Lys | Ala | Ser | Ser | Val | Leu | Ala | Ala |      |
|    |     |     |     | 725 |     |     |     |     |     | 730 |     |     |     |     | 735 |     |      |
| 45 | TTT | GAA | GCT | AAC | TCG | TTA | ACG | GCA | GAA | CAA | CTG | GCT | GAT | GCC | ATG | AAT | 2256 |
|    | Phe | Glu | Ala | Asn | Ser | Leu | Thr | Ala | Glu | Gln | Leu | Ala | Asp | Ala | Met | Asn |      |
|    |     |     |     | 740 |     |     |     |     | 745 |     |     |     | 750 |     |     |     |      |
| 50 | CTT | GAT | GCT | AAT | TTG | CTG | TTG | CAA | GCC | AGT | ATT | CAA | GCA | CAA | AAT | CAT | 2304 |
|    | Leu | Asp | Ala | Asn | Leu | Leu | Leu | Gln | Ala | Ser | Ile | Gln | Ala | Gln | Asn | His |      |
|    |     |     | 755 |     |     |     |     | 760 |     |     |     |     | 765 |     |     |     |      |
| 55 | CAA | CAT | CTT | CCC | CCA | GTA | ACT | CCA | GAA | AAT | GCG | TTC | TCC | TGT | TGG | ACA | 2352 |
|    | Gln | His | Leu | Pro | Pro | Val | Thr | Pro | Glu | Asn | Ala | Phe | Ser | Cys | Trp | Thr |      |
|    |     | 770 |     |     |     |     | 775 |     |     |     |     | 780 |     |     |     |     |      |
| 60 | TCT | ATC | AAT | ACT | ATC | CTG | CAA | TGG | GTT | AAT | GTC | GCA | CAA | CAA | TTG | AAT | 2400 |
|    | Ser | Ile | Asn | Thr | Ile | Leu | Gln | Trp | Val | Asn | Val | Ala | Gln | Gln | Leu | Asn |      |
|    | 785 |     |     |     |     | 790 |     |     |     |     | 795 |     |     |     | 800 |     |      |
| 65 | GTC | GCC | CCA | CAG | GGC | GTT | TCC | GCT | TTG | GTC | GGG | CTG | GAT | TAT | ATT | CAA | 2448 |
|    | Val | Ala | Pro | Gln | Gly | Val | Ser | Ala | Leu | Val | Gly | Leu | Asp | Tyr | Ile | Gln |      |
|    |     |     |     | 805 |     |     |     |     |     | 810 |     |     |     |     | 815 |     |      |
| 70 | TCA | ATG | AAA | GAG | ACA | CCG | ACC | TAT | GCC | CAG | TGG | GAA | AAC | GCG | GCA | GGC | 2496 |
|    | Ser | Met | Lys | Glu | Thr | Pro | Thr | Tyr | Ala | Gln | Trp | Glu | Asn | Ala | Ala | Gly |      |
|    |     |     |     | 820 |     |     |     |     | 825 |     |     |     |     | 830 |     |     |      |
| 75 | GTA | TTA | ACC | GCC | GGG | TTG | AAT | TCA | CAA | CAG | GCT | AAT | ACA | TTA | CAC | GCT | 2544 |
|    | Val | Leu | Thr | Ala | Gly | Leu | Asn | Ser | Gln | Gln | Ala | Asn | Thr | Leu | His | Ala |      |
|    |     |     | 835 |     |     |     |     | 840 |     |     |     |     | 845 |     |     |     |      |
| 80 | TTT | CTG | GAT | GAA | TCT | CGC | AGT | GCC | GCA | TTA | AGC | ACC | TAC | TAT | ATC | CGT | 2592 |
|    | Phe | Leu | Asp | Glu | Ser | Arg | Ser | Ala | Ala | Leu | Ser | Thr | Tyr | Tyr | Ile | Arg |      |
|    |     | 850 |     |     |     |     | 855 |     |     |     |     | 860 |     |     |     |     |      |
| 85 | CAA | GTC | GCC | AAG | GCA | GCG | GCG | GCT | ATT | AAA | AGC | CGT | GAT | GAC | TTG | TAT | 2640 |
|    | Gln | Val | Ala | Lys | Ala | Ala | Ala | Ala | Ile | Lys | Ser | Arg | Asp | Asp | Leu | Tyr |      |
|    | 865 |     |     |     |     | 870 |     |     |     |     | 875 |     |     |     |     | 880 |      |

|    |      |      |      |      |      |      |      |      |      |     |      |      |      |      |      |      |      |
|----|------|------|------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|
|    | CAA  | TAC  | TTA  | CTG  | ATT  | GAT  | AAT  | CAG  | GTT  | TCT | GCG  | GCA  | ATA  | AAA  | ACC  | ACT  | 1888 |
|    | Gln  | Tyr  | Leu  | Leu  | Ile  | Asp  | Asn  | Gln  | Val  | Ser | Ala  | Ala  | Ile  | Lys  | Thr  | Thr  |      |
|    |      |      |      |      | 385  |      |      |      |      | 890 |      |      |      |      | 895  |      |      |
| 5  | CGG  | ATC  | GCC  | GAA  | GCC  | ATT  | GCC  | AGT  | ATT  | CAA | CTG  | TAC  | CTC  | AAC  | CGG  | GCA  | 2735 |
|    | Arg  | Ile  | Ala  | Glu  | Ala  | Ile  | Ala  | Ser  | Ile  | Gln | Leu  | Tyr  | Val  | Asn  | Arg  | Ala  |      |
|    |      |      |      | 900  |      |      |      |      | 905  |     |      |      |      | 910  |      |      |      |
| 10 | TTG  | GAA  | AAT  | GTG  | GAA  | GAA  | AAT  | GCC  | AAT  | TCG | GGG  | GTT  | ATC  | AGC  | CGC  | CAA  | 2784 |
|    | Leu  | Glu  | Asn  | Val  | Glu  | Glu  | Asn  | Ala  | Asn  | Ser | Gly  | Val  | Ile  | Ser  | Arg  | Gln  |      |
|    |      |      | 915  |      |      |      |      | 920  |      |     |      |      | 925  |      |      |      |      |
| 15 | TTC  | TTT  | ATC  | GAC  | TGG  | GAC  | AAA  | TAC  | AAT  | AAA | GCG  | TAC  | AGC  | ACT  | TGG  | GCG  | 2832 |
|    | Phe  | Phe  | Ile  | Asp  | Trp  | Asp  | Lys  | Tyr  | Asn  | Lys | Arg  | Tyr  | Ser  | Thr  | Trp  | Ala  |      |
|    |      | 930  |      |      |      |      | 935  |      |      |     |      | 940  |      |      |      |      |      |
| 20 | GGT  | GTT  | TCT  | CAA  | TTA  | GTT  | TAC  | TAC  | CCG  | GAA | AAC  | TAT  | ATT  | GAT  | CCG  | ACC  | 2880 |
|    | Gly  | Val  | Ser  | Gln  | Leu  | Val  | Tyr  | Tyr  | Pro  | Glu | Asn  | Tyr  | Ile  | Asp  | Pro  | Thr  |      |
|    | 945  |      |      |      |      | 950  |      |      |      |     | 955  |      |      |      |      | 960  |      |
| 25 | ATG  | CGT  | ATC  | GGA  | CAA  | ACC  | AAA  | ATG  | ATG  | GAC | GCA  | TTA  | CTG  | CAA  | TCC  | GTC  | 2928 |
|    | Met  | Arg  | Ile  | Gly  | Gln  | Thr  | Lys  | Met  | Met  | Asp | Ala  | Leu  | Leu  | Gln  | Ser  | Val  |      |
|    |      |      |      | 965  |      |      |      |      |      | 970 |      |      |      |      | 975  |      |      |
| 30 | AGC  | CAA  | AGC  | CAA  | TTA  | AAC  | GCC  | GAT  | ACC  | GTC | GAA  | GAT  | GCC  | TTT  | ATG  | TCT  | 2976 |
|    | Ser  | Gln  | Ser  | Gln  | Leu  | Asn  | Ala  | Asp  | Thr  | Val | Glu  | Asp  | Ala  | Phe  | Met  | Ser  |      |
|    |      |      |      | 980  |      |      |      |      | 985  |     |      |      |      | 990  |      |      |      |
| 35 | TAT  | CTG  | ACA  | TCG  | TTT  | GAA  | CAA  | GTG  | GCT  | AAT | CTT  | AAA  | GTT  | ATT  | AGC  | GCA  | 3024 |
|    | Tyr  | Leu  | Thr  | Ser  | Phe  | Glu  | Gln  | Val  | Ala  | Asn | Leu  | Lys  | Val  | Ile  | Ser  | Ala  |      |
|    |      |      | 995  |      |      |      |      | 1000 |      |     |      |      | 1005 |      |      |      |      |
| 40 | TAT  | CAC  | GAT  | AAT  | ATT  | AAT  | AAC  | GAT  | CAA  | GGG | CTG  | ACC  | TAT  | TTT  | ATC  | GGA  | 3072 |
|    | Tyr  | His  | Asp  | Asn  | Ile  | Asn  | Asn  | Asp  | Gln  | Gly | Leu  | Thr  | Tyr  | Phe  | Ile  | Gly  |      |
|    |      | 1010 |      |      |      |      | 1015 |      |      |     |      | 1020 |      |      |      |      |      |
| 45 | CTC  | AGT  | GAA  | ACT  | GAT  | GCC  | GGT  | GAA  | TAT  | TAT | TGG  | CGC  | AGT  | GTC  | GAT  | CAC  | 3120 |
|    | Leu  | Ser  | Glu  | Thr  | Asp  | Ala  | Gly  | Glu  | Tyr  | Tyr | Trp  | Arg  | Ser  | Val  | Asp  | His  |      |
|    | 1025 |      |      |      |      | 1030 |      |      |      |     | 1035 |      |      |      |      | 1040 |      |
| 50 | AGT  | AAA  | TTC  | AAC  | GAC  | GGT  | AAA  | TTC  | GCG  | GCT | AAT  | GCC  | TGG  | AGT  | GAA  | TGG  | 3168 |
|    | Ser  | Lys  | Phe  | Asn  | Asp  | Gly  | Lys  | Phe  | Ala  | Ala | Asn  | Ala  | Trp  | Ser  | Glu  | Trp  |      |
|    |      |      |      | 1045 |      |      |      |      | 1050 |     |      |      |      | 1055 |      |      |      |
| 55 | CAT  | AAA  | ATT  | GAT  | TGT  | CCA  | ATT  | AAC  | CCT  | TAT | AAA  | AGC  | ACT  | ATC  | CGT  | CCA  | 3216 |
|    | His  | Lys  | Ile  | Asp  | Cys  | Pro  | Ile  | Asn  | Pro  | Tyr | Lys  | Ser  | Thr  | Ile  | Arg  | Pro  |      |
|    |      |      |      | 1060 |      |      |      |      | 1065 |     |      |      |      | 1070 |      |      |      |
| 60 | GTG  | ATA  | TAT  | AAA  | TCC  | CGC  | CTG  | TAT  | CTG  | CTC | TGG  | TTG  | GAA  | CAA  | AAG  | GAG  | 3264 |
|    | Val  | Ile  | Tyr  | Lys  | Ser  | Arg  | Leu  | Tyr  | Leu  | Leu | Trp  | Leu  | Glu  | Gln  | Lys  | Glu  |      |
|    |      |      | 1075 |      |      |      | 1080 |      |      |     |      |      | 1085 |      |      |      |      |
| 65 | ATC  | ACC  | AAA  | CAG  | ACA  | GGA  | AAT  | AGT  | AAA  | GAT | GGC  | TAT  | CAA  | ACT  | GAA  | ACG  | 3312 |
|    | Ile  | Thr  | Lys  | Gln  | Thr  | Gly  | Asn  | Ser  | Lys  | Asp | Gly  | Tyr  | Gln  | Thr  | Glu  | Thr  |      |
|    |      | 1090 |      |      |      |      | 1095 |      |      |     |      | 1100 |      |      |      |      |      |
| 70 | GAT  | TAT  | CGT  | TAT  | GAA  | CTA  | AAA  | TTG  | GCG  | CAT | ATC  | CGC  | TAT  | GAT  | GGC  | ACT  | 3360 |
|    | Asp  | Tyr  | Arg  | Tyr  | Glu  | Leu  | Lys  | Leu  | Ala  | His | Ile  | Arg  | Tyr  | Asp  | Gly  | Thr  |      |
|    | 1105 |      |      |      | 1110 |      |      |      |      |     | 1115 |      |      |      |      | 1120 |      |
| 75 | TGG  | AAT  | ACG  | CCA  | ATC  | ACC  | TTT  | GAT  | GTC  | AAT | AAA  | AAA  | ATA  | TCC  | GAG  | CTA  | 3408 |
|    | Trp  | Asn  | Thr  | Pro  | Ile  | Thr  | Phe  | Asp  | Val  | Asn | Lys  | Lys  | Ile  | Ser  | Glu  | Leu  |      |
|    |      |      |      | 1125 |      |      |      |      | 1130 |     |      |      |      |      | 1135 |      |      |
| 80 | AAA  | CTG  | GAA  | AAA  | AAT  | AGA  | GCG  | CCC  | GGA  | CTC | TAT  | TGT  | GCC  | GGT  | TAT  | CAA  | 3456 |
|    | Lys  | Leu  | Glu  | Lys  | Asn  | Arg  | Ala  | Pro  | Gly  | Leu | Tyr  | Cys  | Ala  | Gly  | Tyr  | Gln  |      |
|    |      |      |      | 1140 |      |      |      | 1145 |      |     |      |      |      | 1150 |      |      |      |
| 85 | GGT  | GAA  | GAT  | ACG  | TTG  | CTG  | GTG  | ATG  | TTT  | TAT | AAC  | CAA  | CAA  | GAC  | ACA  | CTA  | 3504 |
|    | Gly  | Glu  | Asp  | Thr  | Leu  | Leu  | Val  | Met  | Phe  | Tyr | Asn  | Gln  | Gln  | Asp  | Thr  | Leu  |      |
|    |      |      | 1155 |      |      |      |      | 1160 |      |     |      |      |      | 1165 |      |      |      |

|    |                                                                                        |
|----|----------------------------------------------------------------------------------------|
| 5  | GAT AGT TAT AAA AAC GCT TCA ATG CAA GGA CTA TAT ATC TTT GCT GAT 3552                   |
|    | Asp Ser Tyr Lys Asn Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp<br>1170 1175 1180      |
| 10 | ATG GCA TCC AAA GAT ATG ACC CCA GAA CAG AGC AAT GTT TAT CGG GAT 3600                   |
|    | Met Ala Ser Lys Asp Met Thr Pro Glu Gln Ser Asn Val Tyr Arg Asp<br>1185 1190 1195 1200 |
| 15 | AAT AGC TAT CAA CAA TTT GAT ACC AAT AAT GTC AGA AGA GTG AAT AAC 3648                   |
|    | Asn Ser Tyr Gln Gln Phe Asp Thr Asn Asn Val Arg Arg Val Asn Asn<br>1205 1210 1215      |
| 20 | CGC TAT GCA GAG GAT TAT GAG ATT CCT TCC TCG GTA AGT AGC CGT AAA 3696                   |
|    | Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Ser Ser Arg Lys<br>1220 1225 1230      |
| 25 | GAC TAT GGT TGG GGA GAT TAT TAC CTC AGC ATG GTA TAT AAC GGA GAT 3744                   |
|    | Asp Tyr Gly Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp<br>1235 1240 1245      |
| 30 | ATT CCA ACT ATC AAT TAC AAA GCC GCA TCA AGT GAT TTA AAA ATC TAT 3792                   |
|    | Ile Pro Thr Ile Asn Tyr Lys Ala Ala Ser Ser Asp Leu Lys Ile Tyr<br>1250 1255 1260      |
| 35 | ATC TCA CCA AAA TTA AGA ATT ATT CAT AAT GGA TAT GAA GGA CAG AAG 3840                   |
|    | Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Lys<br>1265 1270 1275 1280 |
| 40 | CGC AAT CAA TGC AAT CTG ATG AAT AAA TAT GGC AAA CTA GGT GAT AAA 3888                   |
|    | Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys<br>1285 1290 1295      |
| 45 | TTT ATT GTT TAT ACT AGC TTG GGG GTC AAT CCA AAT AAC TCG TCA AAT 3936                   |
|    | Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn<br>1300 1305 1310      |
| 50 | AAG CTC ATG TTT TAC CCC GTC TAT CAA TAT AGC GGA AAC ACC AGT GGA 3984                   |
|    | Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Ser Gly Asn Thr Ser Gly<br>1315 1320 1325      |
| 55 | CTC AAT CAA GGG AGA CTA CTA TTC CAC CGT GAC ACC ACT TAT CCA TCT 4032                   |
|    | Leu Asn Gln Gly Arg Leu Leu Phe His Arg Asp Thr Thr Pro Ser<br>1330 1335 1340          |
| 60 | AAA GTA GAA GCT TGG ATT CCT GGA GCA AAA CGT TCT CTA ACC AAC CAA 4080                   |
|    | Lys Val Glu Ala Trp Ile Pro Gly Ala Lys Arg Ser Leu Thr Asn Gln<br>1345 1350 1355 1360 |
| 65 | AAT GCC GCC ATT GGT GAT GAT TAT GCT ACA GAC TCT CTG AAT AAA CCG 4128                   |
|    | Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro<br>1365 1370 1375      |
| 70 | GAT GAT CTT AAG CAA TAT ATC TTT ATG ACT GAC AGT AAA GGG ACT GCT 4176                   |
|    | Asp Asp Leu Lys Gln Tyr Ile Phe Met Thr Asp Ser Lys Gly Thr Ala<br>1380 1385 1390      |
| 75 | ACT GAT GTC TCA GGC CCA GTA GAG ATT AAT ACT GCA ATT TCT CCA GCA 4224                   |
|    | Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala Ile Ser Pro Ala<br>1395 1400 1405      |
| 80 | AAA GTT CAG ATA ATA GTC AAA GCG GGT GGC AAG GAG CAA ACT TTT ACC 4272                   |
|    | Lys Val Gln Ile Ile Val Lys Ala Gly Gly Lys Glu Gln Thr Phe Thr<br>1410 1415 1420      |
| 85 | GCA GAT AAA GAT GTC TCC ATT CAG CCA TCA CCT AGC TTT GAT GAA ATG 4320                   |
|    | Ala Asp Lys Asp Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met<br>1425 1430 1435 1440 |
| 90 | AAT TAT CAA TTT AAT GCC CTT GAA ATA GAC GGT TCT GGT CTG AAT TTT 4368                   |
|    | Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Gly Leu Asn Phe<br>1445 1450 1455      |

|    |     |      |      |      |     |      |      |      |     |      |      |      |      |     |      |      |      |
|----|-----|------|------|------|-----|------|------|------|-----|------|------|------|------|-----|------|------|------|
|    |     |      |      | 1445 |     |      |      | 1450 |     |      |      |      | 1455 |     |      |      |      |
| 5  | ATT | AAC  | AAC  | TCA  | GCC | AGT  | ATT  | GAT  | GTT | ACT  | TTT  | ACC  | GCA  | TTT | GCG  | GAG  | 4416 |
|    | Ile | Asn  | Asn  | Ser  | Ala | Ser  | Ile  | Asp  | Val | Thr  | Phe  | Thr  | Ala  | Phe | Ala  | Glu  |      |
|    |     |      |      | 1460 |     |      |      | 1465 |     |      |      |      | 1470 |     |      |      |      |
| 10 | GAT | GCG  | CGC  | AAA  | CTG | GGT  | TAT  | GAA  | AGT | TTC  | AGT  | ATT  | CCT  | GTT | ACC  | CTC  | 4464 |
|    | Asp | Gly  | Arg  | Lys  | Leu | Gly  | Tyr  | Glu  | Ser | Phe  | Ser  | Ile  | Pro  | Val | Thr  | Leu  |      |
|    |     |      | 1475 |      |     |      |      | 1480 |     |      |      |      | 1485 |     |      |      |      |
| 15 | AAG | GTA  | AGT  | ACC  | GAT | AAT  | GCC  | CTG  | ACC | CTG  | CAC  | CAT  | AAT  | GAA | AAT  | GGT  | 4512 |
|    | Lys | Val  | Ser  | Thr  | Asp | Asn  | Ala  | Leu  | Thr | Leu  | His  | His  | Asn  | Glu | Asn  | Gly  |      |
|    |     | 1490 |      |      |     |      | 1495 |      |     |      |      | 1500 |      |     |      |      |      |
| 20 | GCG | CAA  | TAT  | ATG  | CAA | TGG  | CAA  | TCC  | TAT | CGT  | ACC  | CGC  | CTG  | AAT | ACT  | CTA  | 4560 |
|    | Ala | Gln  | Tyr  | Met  | Gln | Trp  | Gln  | Ser  | Tyr | Arg  | Thr  | Arg  | Leu  | Asn | Thr  | Leu  |      |
|    |     | 1505 |      |      |     | 1510 |      |      |     | 1515 |      |      |      |     |      | 1520 |      |
| 25 | TTT | GCC  | CGC  | CAG  | TTG | GTT  | GCA  | CGC  | GCC | ACC  | ACC  | GGA  | ATC  | GAT | ACA  | ATT  | 4608 |
|    | Phe | Ala  | Arg  | Gln  | Leu | Val  | Ala  | Arg  | Ala | Thr  | Thr  | Gly  | Ile  | Asp | Thr  | Ile  |      |
|    |     |      |      | 1525 |     |      |      | 1530 |     |      |      |      | 1535 |     |      |      |      |
| 30 | CTG | AGT  | ATG  | GAA  | ACT | CAG  | AAT  | ATT  | CAG | GAA  | CCG  | CAG  | TTA  | GGC | AAA  | GGT  | 4656 |
|    | Leu | Ser  | Met  | Glu  | Thr | Gln  | Asn  | Ile  | Gln | Glu  | Pro  | Gln  | Leu  | Gly | Lys  | Gly  |      |
|    |     |      |      | 1540 |     |      |      | 1545 |     |      |      |      | 1550 |     |      |      |      |
| 35 | TTC | TAT  | GCT  | ACG  | TTC | GTG  | ATA  | CCT  | CCC | TAT  | AAC  | CTA  | TCA  | ACT | CAT  | GGT  | 4704 |
|    | Phe | Tyr  | Ala  | Thr  | Phe | Val  | Ile  | Pro  | Pro | Tyr  | Asn  | Leu  | Ser  | Thr | His  | Gly  |      |
|    |     |      | 1555 |      |     |      | 1560 |      |     |      |      | 1565 |      |     |      |      |      |
| 40 | GAT | GAA  | CGT  | TGG  | TTT | AAG  | CTT  | TAT  | ATC | AAA  | CAT  | GTT  | GTT  | GAT | AAT  | AAT  | 4752 |
|    | Asp | Glu  | Arg  | Trp  | Phe | Lys  | Leu  | Tyr  | Ile | Lys  | His  | Val  | Val  | Asp | Asn  | Asn  |      |
|    |     | 1570 |      |      |     | 1575 |      |      |     |      | 1580 |      |      |     |      |      |      |
| 45 | TCA | CAT  | ATT  | ATC  | TAT | TCA  | GGC  | CAG  | CTA | ACA  | GAT  | ACA  | AAT  | ATA | AAC  | ATC  | 4800 |
|    | Ser | His  | Ile  | Ile  | Tyr | Ser  | Gly  | Gln  | Leu | Thr  | Asp  | Thr  | Asn  | Ile | Asn  | Ile  |      |
|    |     | 1585 |      |      |     | 1590 |      |      |     | 1595 |      |      |      |     |      | 1600 |      |
| 50 | ACA | TTA  | TTT  | ATT  | CCT | CTT  | GAT  | GAT  | GTC | CCA  | TTG  | AAT  | CAA  | GAT | TAT  | CAC  | 4848 |
|    | Thr | Leu  | Phe  | Ile  | Pro | Leu  | Asp  | Asp  | Val | Pro  | Leu  | Asn  | Gln  | Asp | Tyr  | His  |      |
|    |     |      |      | 1605 |     |      |      | 1610 |     |      |      |      |      |     | 1615 |      |      |
| 55 | GCC | AAG  | GTT  | TAT  | ATG | ACC  | TTC  | AAG  | AAA | TCA  | CCA  | TCA  | GAT  | GGT | ACC  | TGG  | 4896 |
|    | Ala | Lys  | Val  | Tyr  | Met | Thr  | Phe  | Lys  | Lys | Ser  | Pro  | Ser  | Asp  | Gly | Thr  | Trp  |      |
|    |     |      | 1620 |      |     |      |      | 1625 |     |      |      |      | 1630 |     |      |      |      |
| 60 | TGG | GCG  | CCT  | CAC  | TTT | GTT  | AGA  | GAT  | GAT | AAA  | GGA  | ATA  | GTA  | ACA | ATA  | AAC  | 4944 |
|    | Trp | Gly  | Pro  | His  | Phe | Val  | Arg  | Asp  | Asp | Lys  | Gly  | Ile  | Val  | Thr | Ile  | Asn  |      |
|    |     |      | 1635 |      |     |      | 1640 |      |     |      |      | 1645 |      |     |      |      |      |
| 65 | CCT | AAA  | TCC  | ATT  | TTG | ACC  | CAT  | TTT  | GAG | AGC  | GTC  | AAT  | GTC  | CTG | AAT  | AAT  | 4992 |
|    | Pro | Lys  | Ser  | Ile  | Leu | Thr  | His  | Phe  | Glu | Ser  | Val  | Asn  | Val  | Leu | Asn  | Asn  |      |
|    |     | 1650 |      |      |     | 1655 |      |      |     |      | 1660 |      |      |     |      |      |      |
| 70 | ATT | AGT  | AGC  | GAA  | CCA | ATG  | GAT  | TTC  | AGC | GGC  | GCT  | AAC  | AGC  | CTC | TAT  | TTC  | 5040 |
|    | Ile | Ser  | Ser  | Glu  | Pro | Met  | Asp  | Phe  | Ser | Gly  | Ala  | Asn  | Ser  | Leu | Tyr  | Phe  |      |
|    |     | 1665 |      |      |     | 1670 |      |      |     | 1675 |      |      |      |     |      | 1680 |      |
| 75 | TGG | GAA  | CTG  | TTC  | TAC | TAT  | ACC  | CCG  | ATG | CTG  | GTT  | GCT  | CAA  | CGT | TTG  | CTG  | 5088 |
|    | Trp | Glu  | Leu  | Phe  | Tyr | Tyr  | Thr  | Pro  | Met | Leu  | Val  | Ala  | Gln  | Arg | Leu  | Leu  |      |
|    |     |      | 1685 |      |     |      |      | 1690 |     |      |      |      |      |     | 1695 |      |      |
| 80 | CAT | GAA  | CAG  | AAC  | TTC | GAT  | GAA  | GCC  | AAC | CGT  | TGG  | CTG  | AAA  | TAT | GTC  | TGG  | 5136 |
|    | His | Glu  | Gln  | Asn  | Phe | Asp  | Glu  | Ala  | Asn | Arg  | Trp  | Leu  | Lys  | Tyr | Val  | Trp  |      |
|    |     |      | 1700 |      |     |      |      | 1705 |     |      |      | 1710 |      |     |      |      |      |
| 85 | AGT | CCA  | TCC  | GGT  | TAT | ATT  | GTC  | CAC  | GGC | CAG  | ATT  | CAG  | AAC  | TAC | CAG  | TGG  | 5184 |
|    | Ser | Pro  | Ser  | Gly  | Tyr | Ile  | Val  | His  | Gly | Gln  | Ile  | Gln  | Asn  | Tyr | Gln  | Trp  |      |
|    |     |      | 1715 |      |     |      | 1720 |      |     |      |      | 1725 |      |     |      |      |      |
| 90 | AAC | GTC  | CGC  | CCG  | TTA | CTG  | GAA  | GAC  | ACC | AGT  | TGG  | AAC  | AGT  | GAT | CCT  | TTG  | 5232 |

Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu  
 1730 1735 1740  
 5 GAT TCC GTC GAT CCT GAC GCG GTA GCA CAG CAC GAT CCA ATG CAC TAC 5230  
 Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr  
 1745 1750 1755 1760  
 10 AAA GTT TCA ACT TTT ATG CGT ACC TTG GAT CTA TTG ATA GCA CGC GGC 5328  
 Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly  
 1765 1770 1775  
 GAC CAT GCT TAT CGC CAA CTG GAA CGA GAT ACA CTC AAC GAA GCG AAG 5376  
 Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys  
 1780 1785 1790  
 15 ATG TGG TAT ATG CAA GCG CTG CAT CTA TTA GGT GAC AAA CCT TAT CTA 5424  
 Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu  
 1795 1800 1805  
 20 CCG CTG AGT ACG ACA TGG ACT GAT CCA CGA CTA GAC AGA GCC GCG GAT 5472  
 Pro Leu Ser Thr Thr Trp Ser Asp Pro Arg Leu Asp Arg Ala Ala Asp  
 1810 1815 1820  
 25 ATC ACT ACC CAA AAT GCT CAC GAC AGC GCA ATA GTC GCT CTG CGG CAG 5520  
 Ile Thr Thr Gln Asn Ala His Asp Ser Ala Ile Val Ala Leu Arg Gln  
 1825 1830 1835 1840  
 AAT ATA CCT ACA CCG GCA CCT TTA TCA 5547  
 30 Asn Ile Pro Thr Pro Ala Pro Leu Ser  
 1845 1849

(2) INFORMATION FOR SEQ ID NO:49:  
 (i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 1849 amino acids  
 (B) TYPE: amino acids  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49 (TcdAii):  
 45 Features From To Description  
 Peptide 1 1849 TcdAii peptide  
 Fragment 1 12 S2 N-terminus (SEQ ID NO:13)  
 Fragment 196 211 (SEQ ID NO:38)  
 Fragment 466 475 (SEQ ID NO:17)  
 Fragment 993 1004 (SEQ ID NO:23; 12/13)  
 50 Fragment 1297 1312 (SEQ ID NO:18)  
 Fragment 1390 1409 (SEQ ID NO:39)  
 Fragment 1532 1554 (SEQ ID NO:21; 19/23)

55 Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val  
 1 5 10 15  
 Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr Leu Thr  
 20 25 30  
 60 Glu Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser Val Tyr  
 35 40 45  
 Tyr Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu Ser Gln  
 50 55 60  
 65 Gln Asn Met Asp Ile Glu Leu Ser Thr Leu Ser Leu Ser Asn Glu Leu  
 65 70 75 80  
 Leu Leu Glu Ser Ile Lys Thr Glu Ser Lys Leu Glu Asn Tyr Thr Lys

|    | 85                                                              | 90  | 95  |
|----|-----------------------------------------------------------------|-----|-----|
|    | Val Met Glu Met Leu Ser Thr Phe Arg Pro Ser Gly Ala Thr Pro Tyr |     |     |
| 5  | 100                                                             | 105 | 110 |
|    | His Asp Ala Tyr Glu Asn Val Arg Glu Val Ile Gln Leu Gln Asp Pro |     |     |
|    | 115                                                             | 120 | 125 |
| 10 | Gly Leu Glu Gln Leu Asn Ala Ser Pro Ala Ile Ala Gly Leu Met His |     |     |
|    | 130                                                             | 135 | 140 |
|    | Gln Ala Ser Leu Leu Gly Ile Asn Ala Ser Ile Ser Pro Glu Leu Phe |     |     |
|    | 145                                                             | 150 | 155 |
| 15 | Asn Ile Leu Thr Glu Glu Ile Thr Glu Gly Asn Ala Glu Glu Leu Tyr |     |     |
|    | 165                                                             | 170 | 175 |
|    | Lys Lys Asn Phe Gly Asn Ile Glu Pro Ala Ser Leu Ala Met Pro Glu |     |     |
|    | 180                                                             | 185 | 190 |
| 20 | Tyr Leu Lys Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Ser Gln Phe |     |     |
|    | 195                                                             | 200 | 205 |
|    | Ile Gly Lys Ala Ser Asn Phe Gly Gln Gln Glu Tyr Ser Asn Asn Gln |     |     |
| 25 | 210                                                             | 215 | 220 |
|    | Leu Ile Thr Pro Val Val Asn Ser Ser Asp Gly Thr Val Lys Val Tyr |     |     |
|    | 225                                                             | 230 | 235 |
| 30 | Arg Ile Thr Arg Glu Tyr Thr Thr Asn Ala Tyr Gln Met Asp Val Glu |     |     |
|    | 245                                                             | 250 | 255 |
|    | Leu Phe Pro Phe Gly Gly Glu Asn Tyr Arg Leu Asp Tyr Lys Phe Lys |     |     |
|    | 260                                                             | 265 | 270 |
| 35 | Asn Phe Tyr Asn Ala Ser Tyr Leu Ser Ile Lys Leu Asn Asp Lys Arg |     |     |
|    | 275                                                             | 280 | 285 |
|    | Glu Leu Val Arg Thr Glu Gly Ala Pro Gln Val Asn Ile Glu Tyr Ser |     |     |
| 40 | 290                                                             | 295 | 300 |
|    | Ala Asn Ile Thr Leu Asn Thr Ala Asp Ile Ser Gln Pro Phe Glu Ile |     |     |
|    | 305                                                             | 310 | 315 |
| 45 | Gly Leu Thr Arg Val Leu Pro Ser Gly Ser Trp Ala Tyr Ala Ala Ala |     |     |
|    | 325                                                             | 330 | 335 |
|    | Lys Phe Thr Val Glu Glu Tyr Asn Gln Tyr Ser Phe Leu Leu Lys Leu |     |     |
|    | 340                                                             | 345 | 350 |
| 50 | Asn Lys Ala Ile Arg Leu Ser Arg Ala Thr Glu Leu Ser Pro Thr Ile |     |     |
|    | 355                                                             | 360 | 365 |
|    | Leu Glu Gly Ile Val Arg Ser Val Asn Leu Gln Leu Asp Ile Asn Thr |     |     |
| 55 | 370                                                             | 375 | 380 |
|    | Asp Val Leu Gly Lys Val Phe Leu Thr Lys Tyr Tyr Met Gln Arg Tyr |     |     |
|    | 385                                                             | 390 | 395 |
| 60 | Ala Ile His Ala Glu Thr Ala Leu Ile Leu Cys Asn Ala Pro Ile Ser |     |     |
|    | 405                                                             | 410 | 415 |
|    | Gln Arg Ser Tyr Asp Asn Gln Pro Ser Gln Phe Asp Arg Leu Phe Asn |     |     |
|    | 420                                                             | 425 | 430 |
| 65 | Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp Glu Glu Ile |     |     |
|    | 435                                                             | 440 | 445 |
| 70 | Asp Leu Asn Ser Gly Ser Thr Gly Asp Trp Arg Lys Thr Ile Leu Lys |     |     |
|    | 450                                                             | 455 | 460 |



Arg Ala Phe Asn Ile Asp Asp Val Ser Leu Phe Arg Leu Leu Lys Ile  
 465 470 475 480  
 5 Thr Asp His Asp Asn Lys Asp Gly Lys Ile Lys Asn Asn Leu Lys Asn  
 485 490 495  
 Leu Ser Asn Leu Tyr Ile Gly Lys Leu Leu Ala Asp Ile His Gln Leu  
 500 505 510  
 10 Thr Ile Asp Glu Leu Asp Leu Leu Leu Ile Ala Val Gly Glu Gly Lys  
 515 520 525  
 Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Thr Leu Ile Arg  
 530 535 540  
 15 Lys Leu Asn Thr Ile Thr Ser Trp Leu His Thr Gln Lys Trp Ser Val  
 545 550 555 560  
 Phe Gln Leu Phe Ile Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr  
 565 570 575  
 20 Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly  
 580 585 590  
 25 Phe Asp Lys Asp Lys Ala Asp Leu Leu His Val Met Ala Pro Tyr Ile  
 595 600 605  
 Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu  
 610 615 620  
 30 Leu Trp Ala Asp Lys Leu Gln Pro Gly Asp Gly Ala Met Thr Ala Glu  
 625 630 635 640  
 Lys Phe Trp Asp Trp Leu Asn Thr Lys Tyr Thr Pro Gly Ser Ser Glu  
 645 650 655  
 35 Ala Val Glu Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala  
 660 665 670  
 40 Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe  
 675 680 685  
 Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ala Ala Thr Gly Ala  
 690 695 700  
 45 Ala Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala  
 705 710 715 720  
 Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala  
 725 730 735  
 50 Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn  
 740 745 750  
 Leu Asp Ala Asn Leu Leu Leu Gln Ala Ser Ile Gln Ala Gln Asn His  
 755 760 765  
 Gln His Leu Pro Pro Val Thr Pro Glu Asn Ala Phe Ser Cys Trp Thr  
 770 775 780  
 60 Ser Ile Asn Thr Ile Leu Gln Trp Val Asn Val Ala Gln Gln Leu Asn  
 785 790 795 800  
 Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp Tyr Ile Gln  
 805 810 815  
 65 Ser Met Lys Glu Thr Pro Thr Tyr Ala Gln Trp Glu Asn Ala Ala Gly  
 820 825 830  
 Val Leu Thr Ala Gly Leu Asn Ser Gln Gln Ala Asn Thr Leu His Ala  
 835 840 845  
 70

|    |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |
|----|------|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|--|
|    | Phe  | Leu | Asp  | Glu  | Ser  | Arg  | Ser  | Ala  | Ala  | Leu  | Ser  | Thr  | Tyr  | Tyr  | Ile  | Arg  |  |
|    | 850  |     |      |      |      |      | 855  |      |      |      |      | 860  |      |      |      |      |  |
| 5  | Gln  | Val | Ala  | Lys  | Ala  | Ala  | Ala  | Ala  | Ile  | Lys  | Ser  | Arg  | Asp  | Asp  | Leu  | Tyr  |  |
|    | 865  |     |      |      |      | 870  |      |      |      |      | 875  |      |      |      |      | 880  |  |
|    | Gln  | Tyr | Leu  | Leu  | Ile  | Asp  | Asn  | Gln  | Val  | Ser  | Ala  | Ala  | Ile  | Lys  | Thr  | Thr  |  |
|    |      |     |      |      | 885  |      |      |      |      | 890  |      |      |      |      | 895  |      |  |
| 10 | Arg  | Ile | Ala  | Glu  | Ala  | Ile  | Ala  | Ser  | Ile  | Gln  | Leu  | Tyr  | Val  | Asn  | Arg  | Ala  |  |
|    |      |     |      | 900  |      |      |      |      | 905  |      |      |      |      | 910  |      |      |  |
|    | Leu  | Glu | Asn  | Val  | Glu  | Glu  | Asn  | Ala  | Asn  | Ser  | Gly  | Val  | Ile  | Ser  | Arg  | Gln  |  |
|    |      |     | 915  |      |      |      |      | 920  |      |      |      |      | 925  |      |      |      |  |
| 15 | Phe  | Phe | Ile  | Asp  | Trp  | Asp  | Lys  | Tyr  | Asn  | Lys  | Arg  | Tyr  | Ser  | Thr  | Trp  | Ala  |  |
|    | 930  |     |      |      |      |      | 935  |      |      |      |      | 940  |      |      |      |      |  |
| 20 | Gly  | Val | Ser  | Gln  | Leu  | Val  | Tyr  | Tyr  | Pro  | Glu  | Asn  | Tyr  | Ile  | Asp  | Pro  | Thr  |  |
|    | 945  |     |      |      |      | 950  |      |      |      |      | 955  |      |      |      |      | 960  |  |
|    | Met  | Arg | Ile  | Gly  | Gln  | Thr  | Lys  | Met  | Met  | Asp  | Ala  | Leu  | Leu  | Gln  | Ser  | Val  |  |
|    |      |     |      |      | 965  |      |      |      |      | 970  |      |      |      |      | 975  |      |  |
| 25 | Ser  | Gln | Ser  | Gln  | Leu  | Asn  | Ala  | Asp  | Thr  | Val  | Glu  | Asp  | Ala  | Phe  | Met  | Ser  |  |
|    |      |     |      | 980  |      |      |      |      | 985  |      |      |      |      | 990  |      |      |  |
|    | Tyr  | Leu | Thr  | Ser  | Phe  | Glu  | Gln  | Val  | Ala  | Asn  | Leu  | Lys  | Val  | Ile  | Ser  | Ala  |  |
|    |      | 995 |      |      |      |      |      | 1000 |      |      |      |      | 1005 |      |      |      |  |
| 30 | Tyr  | His | Asp  | Asn  | Ile  | Asn  | Asn  | Asp  | Gln  | Gly  | Leu  | Thr  | Tyr  | Phe  | Ile  | Gly  |  |
|    | 1010 |     |      |      |      |      | 1015 |      |      |      |      | 1020 |      |      |      |      |  |
| 35 | Leu  | Ser | Glu  | Thr  | Asp  | Ala  | Gly  | Glu  | Tyr  | Tyr  | Trp  | Arg  | Ser  | Val  | Asp  | His  |  |
|    | 1025 |     |      |      |      | 1030 |      |      |      |      | 1035 |      |      |      |      | 1040 |  |
|    | Ser  | Lys | Phe  | Asn  | Asp  | Gly  | Lys  | Phe  | Ala  | Ala  | Asn  | Ala  | Trp  | Ser  | Glu  | Trp  |  |
|    |      |     |      |      | 1045 |      |      |      |      | 1050 |      |      |      |      | 1055 |      |  |
| 40 | His  | Lys | Ile  | Asp  | Cys  | Pro  | Ile  | Asn  | Pro  | Tyr  | Lys  | Ser  | Thr  | Ile  | Arg  | Pro  |  |
|    |      |     |      | 1060 |      |      |      |      | 1065 |      |      |      |      | 1070 |      |      |  |
|    | Val  | Ile | Tyr  | Lys  | Ser  | Arg  | Leu  | Tyr  | Leu  | Leu  | Trp  | Leu  | Glu  | Gln  | Lys  | Glu  |  |
|    |      |     | 1075 |      |      |      |      | 1080 |      |      |      |      | 1085 |      |      |      |  |
| 45 | Ile  | Thr | Lys  | Gln  | Thr  | Gly  | Asn  | Ser  | Lys  | Asp  | Gly  | Tyr  | Gln  | Thr  | Glu  | Thr  |  |
|    | 1090 |     |      |      |      |      | 1095 |      |      |      |      | 1100 |      |      |      |      |  |
| 50 | Asp  | Tyr | Arg  | Tyr  | Glu  | Leu  | Lys  | Leu  | Ala  | His  | Ile  | Arg  | Tyr  | Asp  | Gly  | Thr  |  |
|    | 1105 |     |      |      |      | 1110 |      |      |      |      | 1115 |      |      |      |      | 1120 |  |
|    | Trp  | Asn | Thr  | Pro  | Ile  | Thr  | Phe  | Asp  | Val  | Asn  | Lys  | Lys  | Ile  | Ser  | Glu  | Leu  |  |
|    |      |     |      |      | 1125 |      |      |      |      | 1130 |      |      |      |      | 1135 |      |  |
| 55 | Lys  | Leu | Glu  | Lys  | Asn  | Arg  | Ala  | Pro  | Gly  | Leu  | Tyr  | Cys  | Ala  | Gly  | Tyr  | Gln  |  |
|    |      |     |      | 1140 |      |      |      |      | 1145 |      |      |      |      | 1150 |      |      |  |
|    | Gly  | Glu | Asp  | Thr  | Leu  | Leu  | Val  | Met  | Phe  | Tyr  | Asn  | Gln  | Gln  | Asp  | Thr  | Leu  |  |
|    |      |     | 1155 |      |      |      |      | 1160 |      |      |      |      | 1165 |      |      |      |  |
| 60 | Asp  | Ser | Tyr  | Lys  | Asn  | Ala  | Ser  | Met  | Gln  | Gly  | Leu  | Tyr  | Ile  | Phe  | Ala  | Asp  |  |
|    | 1170 |     |      |      |      |      | 1175 |      |      |      |      | 1180 |      |      |      |      |  |
| 65 | Met  | Ala | Ser  | Lys  | Asp  | Met  | Thr  | Pro  | Glu  | Gln  | Ser  | Asn  | Val  | Tyr  | Arg  | Asp  |  |
|    | 1185 |     |      |      |      | 1190 |      |      |      |      | 1195 |      |      |      |      | 1200 |  |
|    | Asn  | Ser | Tyr  | Gln  | Gln  | Phe  | Asp  | Thr  | Asn  | Asn  | Val  | Arg  | Arg  | Val  | Asn  | Asn  |  |
|    |      |     |      |      | 1205 |      |      |      |      | 1210 |      |      |      |      | 1215 |      |  |
| 70 | Arg  | Tyr | Ala  | Glu  | Asp  | Tyr  | Glu  | Ile  | Pro  | Ser  | Ser  | Val  | Ser  | Ser  | Arg  | Lys  |  |
|    |      |     |      | 1220 |      |      |      |      | 1225 |      |      |      |      | 1230 |      |      |  |

Asp Tyr Gly Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp  
 1235 1240 1245  
 5 Ile Pro Thr Ile Asn Tyr Lys Ala Ala Ser Ser Asp Leu Lys Ile Tyr  
 1250 1255 1260  
 Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Lys  
 1265 1270 1275 1280  
 10 Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys  
 1285 1290 1295  
 Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn  
 1300 1305 1310  
 15 Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Ser Gly Asn Thr Ser Gly  
 1315 1320 1325  
 Leu Asn Gln Gly Arg Leu Leu Phe His Arg Asp Thr Thr Tyr Pro Ser  
 1330 1335 1340  
 Lys Val Glu Ala Trp Ile Pro Gly Ala Lys Arg Ser Leu Thr Asn Gln  
 1345 1350 1355 1360  
 25 Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro  
 1365 1370 1375  
 Asp Asp Leu Lys Gln Tyr Ile Phe Met Thr Asp Ser Lys Gly Thr Ala  
 1380 1385 1390  
 30 Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala Ile Ser Pro Ala  
 1395 1400 1405  
 Lys Val Gln Ile Ile Val Lys Ala Gly Gly Lys Glu Gln Thr Phe Thr  
 1410 1415 1420  
 Ala Asp Lys Asp Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met  
 1425 1430 1435 1440  
 40 Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Gly Leu Asn Phe  
 1445 1450 1455  
 Ile Asn Asn Ser Ala Ser Ile Asp Val Thr Phe Thr Ala Phe Ala Glu  
 1460 1465 1470  
 Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro Val Thr Leu  
 1475 1480 1485  
 50 Lys Val Ser Thr Asp Asn Ala Leu Thr Leu His His Asn Glu Asn Gly  
 1490 1495 1500  
 Ala Gln Tyr Met Gln Trp Gln Ser Tyr Arg Thr Arg Leu Asn Thr Leu  
 1505 1510 1515 1520  
 55 Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile Asp Thr Ile  
 1525 1530 1535  
 Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly Lys Gly  
 1540 1545 1550  
 Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Leu Ser Thr His Gly  
 1555 1560 1565  
 65 Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val Asp Asn Asn  
 1570 1575 1580  
 Ser His Ile Ile Tyr Ser Gly Gln Leu Thr Asp Thr Asn Ile Asn Ile  
 1585 1590 1595 1600  
 70 Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln Asp Tyr His

1605 1610 1615

Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Gly Thr Trp  
1620 1625 1630

5 Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val Thr Ile Asn  
1635 1640 1645

Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val Leu Asn Asn  
1650 1655 1660

10 Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser Leu Tyr Phe  
1665 1670 1675 1680

15 Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu  
1685 1690 1695

His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys Tyr Val Trp  
1700 1705 1710

20 Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn Tyr Gln Trp  
1715 1720 1725

Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu  
1730 1735 1740

25 Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr  
1745 1750 1755 1760

30 Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly  
1765 1770 1775

Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys  
1780 1785 1790

35 Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu  
1795 1800 1805

Pro Leu Ser Thr Thr Trp Ser Asp Pro Arg Leu Asp Arg Ala Ala Asp  
1810 1815 1820

Ile Thr Thr Gln Asn Ala His Asp Ser Ala Ile Val Ala Leu Arg Gln  
1825 1830 1835 1840

45 Asn Ile Pro Thr Pro Ala Pro Leu Ser  
1845 1849

## (2) INFORMATION FOR SEQ ID NO:50:

## 50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1740 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55

## (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50 (TcdA<sub>iii</sub> coding region):

60

TTG CGC AGC GCT AAT ACC CTG ACT GAT CTC TTC CTG CCG CAA ATC AAT 48  
Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile Asn  
1 5 10 15

65

GAA GTG ATG ATG AAT TAC TGG CAG ACA TTA GCT CAG AGA GTA TAC AAT 96  
Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr Asn  
20 25 30

CTG CGT CAT AAC CTC TCT ATC GAC GGC CAG CCG TTA TAT CTG CCA ATC 144

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Leu | Arg | His | Asn | Leu | Ser | Ile | Asp | Gly | Gln | Pro | Leu | Tyr | Leu | Pro | Ile |     |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |
| 5  | TAT | GCC | ACA | CCG | GCC | GAT | CCG | AAA | GCG | TTA | CTC | AGC | GCC | GCC | GTT | GCC | 192 |
|    | Tyr | Ala | Thr | Pro | Ala | Asp | Pro | Lys | Ala | Leu | Leu | Ser | Ala | Ala | Val | Ala |     |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 50  |     |     |     |     |     |
| 10 | ACT | TCT | CAA | GGT | GGA | GGC | AAG | CTA | CCG | GAA | TCA | TTT | ATG | TCC | CTG | TGG | 240 |
|    | Thr | Ser | Gln | Gly | Gly | Gly | Lys | Leu | Pro | Glu | Ser | Phe | Met | Ser | Leu | Trp |     |
|    |     | 65  |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |     |
| 15 | CGT | TTC | CCG | CAC | ATG | CTG | GAA | AAT | GCG | CCG | GGC | ATG | GTT | AGC | CAG | CTC | 288 |
|    | Arg | Phe | Pro | His | Met | Leu | Glu | Asn | Ala | Arg | Gly | Met | Val | Ser | Gln | Leu |     |
|    |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     |     | 95  |     |     |
| 20 | ACC | CAG | TTC | GGC | TCC | ACG | TTA | CAA | AAT | ATT | ATC | GAA | CGT | CAG | GAC | GCG | 336 |
|    | Thr | Gln | Phe | Gly | Ser | Thr | Leu | Gln | Asn | Ile | Ile | Glu | Arg | Gln | Asp | Ala |     |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |     |
| 25 | GAA | GCG | CTC | AAT | GCG | TTA | TTA | CAA | AAT | CAG | GCC | GCC | GAG | CTG | ATA | TTG | 384 |
|    | Glu | Ala | Leu | Asn | Ala | Leu | Leu | Gln | Asn | Gln | Ala | Ala | Glu | Leu | Ile | Leu |     |
|    |     |     |     | 115 |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
| 30 | ACT | AAC | CTG | AGC | ATT | CAG | GAC | AAA | ACC | ATT | GAA | GAA | TTG | GAT | GCC | GAG | 432 |
|    | Thr | Asn | Leu | Ser | Ile | Gln | Asp | Lys | Thr | Ile | Glu | Glu | Leu | Asp | Ala | Glu |     |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |     |
| 35 | AAA | ACG | GTG | TTG | GAA | AAA | TCC | AAA | GCG | GGA | GCA | CAA | TCG | CGC | TTT | GAT | 480 |
|    | Lys | Thr | Val | Leu | Glu | Lys | Ser | Lys | Ala | Gly | Ala | Gln | Ser | Arg | Phe | Asp |     |
|    |     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |     |
| 40 | AGC | TAC | GGC | AAA | CTG | TAC | GAT | GAG | AAT | ATC | AAC | GCC | GGT | GAA | AAC | CAA | 528 |
|    | Ser | Tyr | Gly | Lys | Leu | Tyr | Asp | Glu | Asn | Ile | Asn | Ala | Gly | Glu | Asn | Gln |     |
|    |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |     |
| 45 | GCC | ATG | ACG | CTA | CGA | GCG | TCC | GCC | GCC | GGG | CTT | ACC | ACG | GCA | GTT | CAG | 576 |
|    | Ala | Met | Thr | Leu | Arg | Ala | Ser | Ala | Ala | Gly | Leu | Thr | Thr | Ala | Val | Gln |     |
|    |     |     |     | 180 |     |     |     | 185 |     |     |     |     |     | 190 |     |     |     |
| 50 | GCA | TCC | CGT | CTG | GCC | GGT | GCG | GCG | GCT | GAT | CTG | GTG | CCT | AAC | ATC | TTC | 624 |
|    | Ala | Ser | Arg | Leu | Ala | Gly | Ala | Ala | Ala | Asp | Leu | Val | Pro | Asn | Ile | Phe |     |
|    |     |     | 195 |     |     |     | 200 |     |     |     |     |     | 205 |     |     |     |     |
| 55 | GGC | TTT | GCC | GGT | GGC | GGC | AGC | CGT | TGG | GGG | GCT | ATC | GCT | GAG | GCG | ACA | 672 |
|    | Gly | Phe | Ala | Gly | Gly | Gly | Ser | Arg | Trp | Gly | Ala | Ile | Ala | Glu | Ala | Thr |     |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |     |
| 60 | GGT | TAT | GTG | ATG | GAA | TTC | TCC | GCG | AAT | GTT | ATG | AAC | ACC | GAA | GCG | GAT | 720 |
|    | Gly | Tyr | Val | Met | Glu | Phe | Ser | Ala | Asn | Val | Met | Asn | Thr | Glu | Ala | Asp |     |
|    |     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |     |
| 65 | AAA | ATT | AGC | CAA | TCT | GAA | ACC | TAC | CGT | CGT | CGC | CGT | CAG | GAG | TGG | GAG | 768 |
|    | Lys | Ile | Ser | Gln | Ser | Glu | Thr | Tyr | Arg | Arg | Arg | Arg | Gln | Glu | Trp | Glu |     |
|    |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |     |
| 70 | ATC | CAG | CGG | AAT | AAT | GCC | GAA | GCG | GAA | TTG | AAG | CAA | ATC | GAT | GCT | CAG | 816 |
|    | Ile | Gln | Arg | Asn | Asn | Ala | Glu | Ala | Glu | Leu | Lys | Gln | Ile | Asp | Ala | Gln |     |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |     |
| 75 | CTC | AAA | TCA | CTC | GCT | GTA | CGC | CGC | GAA | GCC | GCC | GTA | TTG | CAG | AAA | ACC | 864 |
|    | Leu | Lys | Ser | Leu | Ala | Val | Arg | Arg | Glu | Ala | Ala | Val | Leu | Gln | Lys | Thr |     |
|    |     |     | 275 |     |     |     | 280 |     |     |     |     |     | 285 |     |     |     |     |
| 80 | AGT | CTG | AAA | ACC | CAA | CAA | GAA | CAG | ACC | CAA | TCT | CAA | TTG | GCC | TTC | CTG | 912 |
|    | Ser | Leu | Lys | Thr | Gln | Gln | Glu | Gln | Thr | Gln | Ser | Gln | Leu | Ala | Phe | Leu |     |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |     |
| 85 | CAA | CGT | AAG | TTC | AGC | AAT | CAG | GCG | TTA | TAC | AAC | TGG | CTG | CGT | GGT | CGA | 960 |
|    | Gln | Arg | Lys | Phe | Ser | Asn | Gln | Ala | Leu | Tyr | Asn | Trp | Leu | Arg | Gly | Arg |     |
|    |     | 305 |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |     |

|    |     |             |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | CTG | SCG         | GCG | ATT | TAC | TTC | CAG | TTC | TAC | GAT | TTG | GCC | STC | GCG | CGT | TGC | 1008 |
|    | Leu | Ala         | Ala | Ile | Tyr | Phe | Gln | Phe | Tyr | Asp | Leu | Ala | Val | Ala | Arg | Cys |      |
|    |     |             |     | 325 |     |     |     |     |     | 330 |     |     |     |     | 335 |     |      |
| 5  | CTG | ATG         | GCA | GAA | CAA | GCT | TAC | CGT | TGG | GAA | CTC | AAT | GAT | GAC | TCT | GCC | 1056 |
|    | Leu | Met         | Ala | Glu | Gln | Ala | Tyr | Arg | Trp | Glu | Leu | Asn | Asp | Asp | Ser | Ala |      |
|    |     |             |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |      |
| 10 | CGC | TTC         | ATT | AAA | CCG | GGC | GCC | TGG | CAG | GGA | ACC | TAT | GCC | GGT | CTG | CTT | 1104 |
|    | Arg | Phe         | Ile | Lys | Pro | Gly | Ala | Trp | Gln | Gly | Thr | Tyr | Ala | Gly | Leu | Leu |      |
|    |     |             | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |      |
| 15 | GCA | GGT         | GAA | ACC | TTG | ATG | CTG | AGT | CTG | GCA | CAA | ATG | GAA | GAC | GCT | CAT | 1152 |
|    | Ala | Gly         | Glu | Thr | Leu | Met | Leu | Ser | Leu | Ala | Gln | Met | Glu | Asp | Ala | His |      |
|    |     | 370         |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |      |
| 20 | CTG | AAA         | CGC | GAT | AAA | CGC | GCA | TTA | GAG | GTT | GAA | CGC | ACA | GTA | TCG | CTG | 1200 |
|    | Leu | Lys         | Arg | Asp | Lys | Arg | Ala | Leu | Glu | Val | Glu | Arg | Thr | Val | Ser | Leu |      |
|    | 385 |             |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |      |
|    | GCC | GAA         | GTT | TAT | GCA | GGA | TTA | CCA | AAA | GAT | AAC | GGT | CCA | TTT | TCC | CTG | 1248 |
|    | Ala | Glu         | Val | Tyr | Ala | Gly | Leu | Pro | Lys | Asp | Asn | Gly | Pro | Phe | Ser | Leu |      |
|    |     |             |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |      |
| 25 | GCT | CAG         | GAA | ATT | GAC | AAG | CTG | GTG | AGT | CAA | GGT | TCA | GGC | AGT | GCC | GGC | 1296 |
|    | Ala | Gln         | Glu | Ile | Asp | Lys | Leu | Val | Ser | Gln | Gly | Ser | Gly | Ser | Ala | Gly |      |
|    |     |             |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |      |
| 30 | AGT | GGT         | AAT | AAT | AAT | TTG | GCG | TTC | GGC | GCC | GGC | ACG | GAC | ACT | AAA | ACC | 1344 |
|    | Ser | Gly         | Asn | Asn | Asn | Leu | Ala | Phe | Gly | Ala | Gly | Thr | Asp | Thr | Lys | Thr |      |
|    |     |             | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |      |
| 35 | TCT | TTG         | CAG | GCA | TCA | GTT | TCA | TTC | GCT | GAT | TTG | AAA | ATT | CGT | GAA | GAT | 1392 |
|    | Ser | Leu         | Gln | Ala | Ser | Val | Ser | Phe | Ala | Asp | Leu | Lys | Ile | Arg | Glu | Asp |      |
|    |     | 450         |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |      |
| 40 | TAC | CCG         | GCA | TCG | CTT | GGC | AAA | ATT | CGA | CGT | ATC | AAA | CAG | ATC | AGC | GTC | 1440 |
|    | Tyr | Pro         | Ala | Ser | Leu | Gly | Lys | Ile | Arg | Arg | Ile | Lys | Gln | Ile | Ser | Val |      |
|    | 465 |             |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |      |
|    | ACT | TTG         | CCC | GCG | CTA | CTG | GGA | CCG | TAT | CAG | GAT | GTA | CAG | GCA | ATA | TTG | 1488 |
|    | Thr | Leu         | Pro | Ala | Leu | Leu | Gly | Pro | Tyr | Gln | Asp | Val | Gln | Ala | Ile | Leu |      |
|    |     |             |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |      |
| 45 | TCT | TAC         | GGC | GAT | AAA | GCC | GGA | TTA | GCT | AAC | GGC | TGT | GAA | GCG | CTG | GCA | 1536 |
|    | Ser | Tyr         | Gly | Asp | Lys | Ala | Gly | Leu | Ala | Asn | Gly | Cys | Glu | Ala | Leu | Ala |      |
|    |     |             |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |      |
| 50 | GTT | TCT         | CAC | GGT | ATG | AAT | GAC | AGC | GGC | CAA | TTC | CAG | CTC | GAT | TTC | AAC | 1584 |
|    | Val | Ser         | His | Gly | Met | Asn | Asp | Ser | Gly | Gln | Phe | Gln | Leu | Asp | Phe | Asn |      |
|    |     |             | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |      |
| 55 | GAT | GGC         | AAA | TTC | CTG | CCA | TTC | GAA | GGC | ATC | GCC | ATT | GAT | CAA | GGC | ACG | 1632 |
|    | Asp | Gly         | Lys | Phe | Leu | Pro | Phe | Glu | Gly | Ile | Ala | Ile | Asp | Gln | Gly | Thr |      |
|    |     | 530         |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |      |
| 60 | CTG | ACA         | CTG | AGC | TTC | CCA | AAT | GCA | TCT | ATG | CCG | GAG | AAA | GGT | AAA | CAA | 1680 |
|    | Leu | Thr         | Leu | Ser | Phe | Pro | Asn | Ala | Ser | Met | Pro | Glu | Lys | Gly | Lys | Gln |      |
|    | 545 |             |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |      |
|    | GCC | ACT         | ATG | TTA | AAA | ACC | CTG | AAC | GAT | ATC | ATT | TTG | CAT | ATT | CGC | TAC | 1728 |
|    | Ala | Thr         | Met | Leu | Lys | Thr | Leu | Asn | Asp | Ile | Ile | Leu | His | Ile | Arg | Tyr |      |
|    |     |             |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |      |
| 65 | ACC | ATT         | AAA | TAA |     |     |     |     |     |     |     |     |     |     |     |     | 1740 |
|    | Thr | Ile         | Lys | ... |     |     |     |     |     |     |     |     |     |     |     |     | 579  |
| 70 | (2) | INFORMATION | FOR | SEQ | ID  | NO: | 51: |     |     |     |     |     |     |     |     |     |      |

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 579 amino acids  
 (B) TYPE: amino acids  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51 (TcdA<sub>iii</sub>):

10

Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile Asn  
 1 5 10 15

15

Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr Asn  
 20 25 30

Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro Ile  
 35 40 45

20

Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ala  
 50 55 60

Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu Trp  
 65 70 75 80

25

Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln Leu  
 85 90 95

30

Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp Ala  
 100 105 110

Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile Leu  
 115 120 125

35

Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala Glu  
 130 135 140

Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe Asp  
 145 150 155 160

40

Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Gly Glu Asn Gln  
 165 170 175

45

Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val Gln  
 180 185 190

Ala Ser Arg Leu Ala Gly Ala Ala Ala Asp Leu Val Pro Asn Ile Phe  
 195 200 205

50

Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala Thr  
 210 215 220

Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala Asp  
 225 230 235 240

55

Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp Glu  
 245 250 255

60

Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala Gln  
 260 265 270

Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys Thr  
 275 280 285

65

Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe Leu  
 290 295 300

Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly Arg  
 305 310 315 320

70

Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg Cys  
 325 330 335  
 5 Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser Ala  
 340 345 350  
 Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu  
 355 360 365  
 10 Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala His  
 370 375 380  
 Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu  
 385 390 395 400  
 15 Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser Leu  
 405 410 415  
 Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala Gly  
 420 425 430  
 20 Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys Thr  
 435 440 445  
 25 Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu Asp  
 450 455 460  
 Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser Val  
 465 470 475 480  
 30 Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile Leu  
 485 490 495  
 Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu Ala  
 500 505 510  
 35 Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn  
 515 520 525  
 40 Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly Thr  
 530 535 540  
 Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys Gln  
 545 550 555 560  
 45 Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg Tyr  
 565 570 575  
 50 Thr Ile Lys ...  
 579

## (2) INFORMATION FOR SEQ ID NO:52:

## 55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5532 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## 60 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52 (TcdA<sub>iii</sub> coding region):

65 TTT ATA CAA GGT TAT AGT GAT CTG TTT GGT AAT CGT GCT GAT AAC TAT 48  
 Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala Asp Asn Tyr  
 1 5 10 15

GCC GCG CCG GGC TCG GTT GCA TCG ATG TTC TCA CCG GCG GCT TAT TTG 96



|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Ala | Ala | Pro | Gly | Ser | Val | Ala | Ser | Met | Phe | Ser | Pro | Ala | Ala | Tyr | Leu |     |
|    |     |     | 20  |     |     |     |     |     | 25  |     |     |     | 30  |     |     |     |     |
| 5  | ACG | GAA | TTG | TAC | CGT | GAA | GCC | AAA | AAC | TTG | CAT | GAC | AGC | AGC | TCA | ATT | 144 |
|    | Thr | Glu | Leu | Tyr | Arg | Glu | Ala | Lys | Asn | Leu | His | Asp | Ser | Ser | Ser | Ile |     |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     | 45  |     |     |     |     |     |
| 10 | TAT | TAC | CTA | GAT | AAA | CGT | CGC | CCG | GAT | TTA | GCA | AGC | TTA | ATG | CTC | AGC | 192 |
|    | Tyr | Tyr | Leu | Asp | Lys | Arg | Arg | Pro | Asp | Leu | Ala | Ser | Leu | Met | Leu | Ser |     |
|    |     |     | 50  |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |     |
| 15 | CAG | AAA | AAT | ATG | GAT | GAG | GAA | ATT | TCA | ACG | CTG | GCT | CTC | TCT | AAT | GAA | 240 |
|    | Gln | Lys | Asn | Met | Asp | Glu | Glu | Ile | Ser | Thr | Leu | Ala | Leu | Ser | Asn | Glu |     |
|    | 65  |     |     |     |     | 70  |     |     |     | 75  |     |     |     |     |     | 80  |     |
|    | TTG | TGC | CTT | GCC | GGG | ATC | GAA | ACA | AAA | ACA | GGA | AAA | TCA | CAA | GAT | GAA | 288 |
|    | Leu | Cys | Leu | Ala | Gly | Ile | Glu | Thr | Lys | Thr | Gly | Lys | Ser | Gln | Asp | Glu |     |
|    |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     |     | 95  |     |     |
| 20 | GTG | ATG | GAT | ATG | TTG | TCA | ACT | TAT | CGT | TTA | AGT | GGA | GAG | ACA | CCT | TAT | 336 |
|    | Val | Met | Asp | Met | Leu | Ser | Thr | Tyr | Arg | Leu | Ser | Gly | Glu | Thr | Pro | Tyr |     |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |     |
| 25 | CAT | CAC | GCT | TAT | GAA | ACT | GTT | CGT | GAA | ATC | GTT | CAT | GAA | CGT | GAT | CCA | 384 |
|    | His | His | Ala | Tyr | Glu | Thr | Val | Arg | Glu | Ile | Val | His | Glu | Arg | Asp | Pro |     |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
| 30 | GGA | TTT | CGT | CAT | TTG | TCA | CAG | GCA | CCC | ATT | GTT | GCT | GCT | AAG | CTC | GAT | 432 |
|    | Gly | Phe | Arg | His | Leu | Ser | Gln | Ala | Pro | Ile | Val | Ala | Ala | Lys | Leu | Asp |     |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |     |
| 35 | CCT | GTG | ACT | TTG | TTG | GGT | ATT | AGC | TCC | CAT | ATT | TCG | CCA | GAA | CTG | TAT | 480 |
|    | Pro | Val | Thr | Leu | Leu | Gly | Ile | Ser | Ser | His | Ile | Ser | Pro | Glu | Leu | Tyr |     |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |     |
|    | AAC | TTG | CTG | ATT | GAG | GAG | ATC | CCG | GAA | AAA | GAT | GAA | GCC | GCG | CTT | GAT | 528 |
|    | Asn | Leu | Leu | Ile | Glu | Glu | Ile | Pro | Glu | Lys | Asp | Glu | Ala | Ala | Leu | Asp |     |
|    |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |     |
| 40 | ACG | CTT | TAT | AAA | ACA | AAC | TTT | GGC | GAT | ATT | ACT | ACT | GCT | CAG | TTA | ATG | 576 |
|    | Thr | Leu | Tyr | Lys | Thr | Asn | Phe | Gly | Asp | Ile | Thr | Thr | Ala | Gln | Leu | Met |     |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |
| 45 | TCC | CCA | AGT | TAT | CTG | GCC | CGG | TAT | TAT | GGC | GTC | TCA | CCG | GAA | GAT | ATT | 624 |
|    | Ser | Pro | Ser | Tyr | Leu | Ala | Arg | Tyr | Tyr | Gly | Val | Ser | Pro | Glu | Asp | Ile |     |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |
| 50 | GCC | TAC | GTG | ACG | ACT | TCA | TTA | TCA | CAT | GTT | GGA | TAT | AGC | AGT | GAT | ATT | 672 |
|    | Ala | Tyr | Val | Thr | Thr | Ser | Leu | Ser | His | Val | Gly | Tyr | Ser | Ser | Asp | Ile |     |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |     |
| 55 | CTG | GTT | ATT | CCG | TTG | GTC | GAT | GGT | GTG | GGT | AAG | ATG | GAA | GTA | GTT | CGT | 720 |
|    | Leu | Val | Ile | Pro | Leu | Val | Asp | Gly | Val | Gly | Lys | Met | Glu | Val | Val | Arg |     |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |     |
|    | GTT | ACC | CGA | ACA | CCA | TCG | GAT | AAT | TAT | ACC | AGT | CAG | ACG | AAT | TAT | ATT | 768 |
|    | Val | Thr | Arg | Thr | Pro | Ser | Asp | Asn | Tyr | Thr | Ser | Gln | Thr | Asn | Tyr | Ile |     |
|    |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |     |
| 60 | GAG | CTG | TAT | CCA | CAG | GGT | GGC | GAC | AAT | TAT | TTG | ATC | AAA | TAC | AAT | CTA | 816 |
|    | Glu | Leu | Tyr | Pro | Gln | Gly | Gly | Asp | Asn | Tyr | Leu | Ile | Lys | Tyr | Asn | Leu |     |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |     |
| 65 | AGC | AAT | AGT | TTT | GGT | TTG | GAT | GAT | TTT | TAT | CTG | CAA | TAT | AAA | GAT | GGT | 864 |
|    | Ser | Asn | Ser | Phe | Gly | Leu | Asp | Asp | Phe | Tyr | Leu | Gln | Tyr | Lys | Asp | Gly |     |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |     |
| 70 | TCC | GCT | GAT | TGG | ACT | GAG | ATT | GCC | CAT | AAT | CCC | TAT | CCT | GAT | ATG | GTC | 912 |
|    | Ser | Ala | Asp | Trp | Thr | Glu | Ile | Gla | Ala | His | Asn | Pro | Pro | Asp | Met | Val |     |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |     |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | ATA | AAT | CAA | AAG | TAT | GAA | TCA | CAG | GCG | ACA | ATC | AAA | CGT | AGT | GAC | TCT | 360  |
|    | Ile | Asn | Gln | Lys | Tyr | Glu | Ser | Gln | Ala | Thr | Ile | Lys | Arg | Ser | Asp | Ser | 320  |
|    | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     |     |      |
| 5  | GAC | AAT | ATA | CTC | AGT | ATA | GGG | TTA | CAA | AGA | TGG | CAT | AGC | GGT | AGT | TAT | 1008 |
|    | Asp | Asn | Ile | Leu | Ser | Ile | Gly | Leu | Gln | Arg | Trp | His | Ser | Gly | Ser | Tyr |      |
|    |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |      |
| 10 | AAT | TTT | GCC | GCC | GCC | AAT | TTT | AAA | ATT | GAC | CAA | TAC | TCC | CCG | AAA | GCT | 1056 |
|    | Asn | Phe | Ala | Ala | Ala | Asn | Phe | Lys | Ile | Asp | Gln | Tyr | Ser | Pro | Lys | Ala |      |
|    |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |      |
| 15 | TTC | CTG | CTT | AAA | ATG | AAT | AAG | GCT | ATT | CGG | TTG | CTC | AAA | GCT | ACC | GGC | 1104 |
|    | Phe | Leu | Leu | Lys | Met | Asn | Lys | Ala | Ile | Arg | Leu | Leu | Lys | Ala | Thr | Gly |      |
|    |     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |      |
| 20 | CTC | TCT | TTT | GCT | ACG | TTG | GAG | CGT | ATT | GTT | GAT | AGT | GTT | AAT | AGC | ACC | 1152 |
|    | Leu | Ser | Phe | Ala | Thr | Leu | Glu | Arg | Ile | Val | Asp | Ser | Val | Asn | Ser | Thr |      |
|    |     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |      |
| 25 | AAA | TCC | ATC | ACG | GTT | GAG | GTA | TTA | AAC | AAG | GTT | TAT | CGG | GTA | AAA | TTC | 1200 |
|    | Lys | Ser | Ile | Thr | Val | Glu | Val | Leu | Asn | Lys | Val | Tyr | Arg | Val | Lys | Phe |      |
|    | 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |      |
| 30 | TAT | ATT | GAT | CGT | TAT | GGC | ATC | AGT | GAA | GAG | ACA | GCC | GCT | ATT | TTG | GCT | 1248 |
|    | Tyr | Ile | Asp | Arg | Tyr | Gly | Ile | Ser | Glu | Glu | Thr | Ala | Ala | Ile | Leu | Ala |      |
|    |     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |      |
| 35 | AAT | ATT | AAT | ATC | TCT | CAG | CAA | GCT | GTT | GGC | AAT | CAG | CTT | AGC | CAG | TTT | 1296 |
|    | Asn | Ile | Asn | Ile | Ser | Gln | Gln | Ala | Val | Gly | Asn | Gln | Leu | Ser | Gln | Phe |      |
|    |     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |      |
| 40 | GAG | CAA | CTA | TTT | AAT | CAC | CCG | CCG | CTC | AAT | GGT | ATT | CGC | TAT | GAA | ATC | 1344 |
|    | Glu | Gln | Leu | Phe | Asn | His | Pro | Pro | Leu | Asn | Gly | Ile | Arg | Tyr | Glu | Ile |      |
|    |     |     | 435 |     |     |     | 440 |     |     |     |     |     | 445 |     |     |     |      |
| 45 | AGT | GAG | GAC | AAC | TCC | AAA | CAT | CTT | CCT | AAT | CCT | GAT | CTG | AAC | CTT | AAA | 1392 |
|    | Ser | Glu | Asp | Asn | Ser | Lys | His | Leu | Pro | Asn | Pro | Asp | Leu | Asn | Leu | Lys |      |
|    |     | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |      |
| 50 | CCA | GAC | AGT | ACC | GGT | GAT | GAT | CAA | CGC | AAG | GCG | GTT | TTA | AAA | CGC | GCG | 1440 |
|    | Pro | Asp | Ser | Thr | Gly | Asp | Asp | Gln | Arg | Lys | Ala | Val | Leu | Lys | Arg | Ala |      |
|    | 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |      |
| 55 | TTT | CAG | GTT | AAC | GCC | AGT | GAG | TTG | TAT | CAG | ATG | TTA | TTG | ATC | ACT | GAT | 1488 |
|    | Phe | Gln | Val | Asn | Ala | Ser | Glu | Leu | Tyr | Gln | Met | Leu | Leu | Ile | Thr | Asp |      |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |      |
| 60 | CGT | AAA | GAA | GAC | GGT | GTT | ATC | AAA | AAT | AAC | TTA | GAG | AAT | TTG | TCT | GAT | 1536 |
|    | Arg | Lys | Glu | Asp | Gly | Val | Ile | Lys | Asn | Asn | Leu | Glu | Asn | Ser | Asp |     |      |
|    |     |     |     | 500 |     |     |     | 505 |     |     |     |     |     | 510 |     |     |      |
| 65 | CTG | TAT | TTG | GTT | AGT | TTG | CTG | GCC | CAG | ATT | CAT | AAC | CTG | ACT | ATT | GCT | 1584 |
|    | Leu | Tyr | Leu | Val | Ser | Leu | Leu | Ala | Gln | Ile | His | Asn | Leu | Thr | Ile | Ala |      |
|    |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |      |
| 70 | GAA | TTG | AAC | ATT | TTG | TTG | GTG | ATT | TGT | GGC | TAT | GGC | GAC | ACC | AAC | ATT | 1632 |
|    | Glu | Leu | Asn | Ile | Leu | Leu | Val | Ile | Cys | Gly | Tyr | Gly | Asp | Thr | Asn | Ile |      |
|    |     | 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |      |
| 75 | TAT | CAG | ATT | ACC | GAC | GAT | AAT | TTA | GCC | AAA | ATA | GTG | GAA | ACA | TTG | TTG | 1680 |
|    | Tyr | Gln | Ile | Thr | Asp | Asp | Asn | Leu | Ala | Lys | Ile | Val | Glu | Thr | Leu | Leu |      |
|    | 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |      |
| 80 | TGG | ATC | ACT | CAA | TGG | TTG | AAG | ACC | CAA | AAA | TGG | ACA | GTT | ACC | GAC | CTG | 1728 |
|    | Trp | Ile | Thr | Gln | Trp | Leu | Lys | Thr | Gln | Lys | Trp | Thr | Val | Thr | Asp | Leu |      |
|    |     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |      |
| 85 | TTT | CTG | ATG | ACC | ACG | GCC | ACT | TAC | AGC | ACC | ACT | TTA | ACG | CCA | GAA | ATT | 1776 |
|    | Phe | Leu | Met | Thr | Thr | Ala | Thr | Tyr | Ser | Thr | Thr | Leu | Thr | Pro | Glu | Ile |      |
|    |     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |      |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | AGC | AAT | CTG | ACG | GCT | ACG | TTG | TCT | TCA | ACT | TTG | CAT | GGC | AAA | GAG | AGT | 1824 |
|    | Ser | Asn | Leu | Thr | Ala | Thr | Leu | Ser | Ser | Thr | Leu | His | Gly | Lys | Glu | Ser |      |
| 5  |     |     | 595 |     |     |     | 600 |     |     |     |     |     | 605 |     |     |     |      |
|    | CTG | ATT | GGG | GAA | GAT | CTG | AAA | AGA | GCA | ATG | GCG | CCT | TGC | TTC | ACT | TCG | 1872 |
|    | Leu | Ile | Gly | Glu | Asp | Leu | Lys | Arg | Ala | Met | Ala | Pro | Cys | Phe | Thr | Ser |      |
|    | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |     |      |
| 10 | GCT | TTG | CAT | TTG | ACT | TCT | CAA | GAA | GTT | GCG | TAT | GAC | CTG | CTG | TTG | TGG | 1920 |
|    | Ala | Leu | His | Leu | Thr | Ser | Gln | Glu | Val | Ala | Tyr | Asp | Leu | Leu | Leu | Trp |      |
|    | 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |      |
| 15 | ATA | GAC | CAG | ATT | CAA | CCG | GCA | CAA | ATA | ACT | GTT | GAT | GGG | TTT | TGG | GAA | 1968 |
|    | Ile | Asp | Gln | Ile | Gln | Pro | Ala | Gln | Ile | Thr | Val | Asp | Gly | Phe | Trp | Glu |      |
|    |     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |      |
| 20 | GAA | GTG | CAA | ACA | ACA | CCA | ACC | AGC | TTG | AAG | GTG | ATT | ACC | TTT | GCT | CAG | 2016 |
|    | Glu | Val | Gln | Thr | Thr | Pro | Thr | Ser | Leu | Lys | Val | Ile | Thr | Phe | Ala | Gln |      |
|    |     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |      |
| 25 | GTG | CTG | GCA | CAA | TTG | AGC | CTG | ATC | TAT | CGT | CGT | ATT | GGG | TTA | AGT | GAA | 2064 |
|    | Val | Leu | Ala | Gln | Leu | Ser | Leu | Ile | Tyr | Arg | Arg | Ile | Gly | Leu | Ser | Glu |      |
|    | 675 |     |     |     |     | 680 |     |     |     |     |     |     | 685 |     |     |     |      |
| 30 | ACG | GAA | CTG | TCA | CTG | ATC | GTG | ACT | CAA | TCT | TCT | CTG | CTA | GTG | GCA | GGC | 2112 |
|    | Thr | Glu | Leu | Ser | Leu | Ile | Val | Thr | Gln | Ser | Ser | Leu | Leu | Val | Ala | Gly |      |
|    | 690 |     |     |     |     | 695 |     |     |     |     |     | 700 |     |     |     |     |      |
| 35 | AAA | AGC | ATA | CTG | GAT | CAC | GGT | CTG | TTA | ACC | CTG | ATG | GCC | TTG | GAA | GGT | 2160 |
|    | Lys | Ser | Ile | Leu | Asp | His | Gly | Leu | Leu | Thr | Leu | Met | Ala | Leu | Glu | Gly |      |
|    | 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     | 720 |      |
| 40 | TTT | CAT | ACC | TGG | GTT | AAT | GGC | TTG | GGG | CAA | CAT | GCC | TCC | TTG | ATA | TTG | 2208 |
|    | Phe | His | Thr | Trp | Val | Asn | Gly | Leu | Gly | Gln | His | Ala | Ser | Leu | Ile | Leu |      |
|    |     |     |     |     | 725 |     |     |     | 730 |     |     |     |     |     | 735 |     |      |
| 45 | GCG | GCG | TTG | AAA | GAC | GGA | GCC | TTG | ACA | GTT | ACC | GAT | GTA | GCA | CAA | GCT | 2256 |
|    | Ala | Ala | Leu | Lys | Asp | Gly | Ala | Leu | Thr | Val | Thr | Asp | Val | Ala | Gln | Ala |      |
|    |     |     |     | 740 |     |     |     |     | 745 |     |     |     |     | 750 |     |     |      |
| 50 | ATG | AAT | AAG | GAG | GAA | TCT | CTC | CTA | CAA | ATG | GCA | GCT | AAT | CAG | GTG | GAG | 2304 |
|    | Met | Asn | Lys | Glu | Glu | Ser | Leu | Leu | Gln | Met | Ala | Ala | Asn | Gln | Val | Glu |      |
|    |     |     | 755 |     |     |     |     | 760 |     |     |     |     | 765 |     |     |     |      |
| 55 | AAG | GAT | CTA | ACA | AAA | CTG | ACC | AGT | TGG | ACA | CAG | ATT | GAC | GCT | ATT | CTG | 2352 |
|    | Lys | Asp | Leu | Thr | Lys | Leu | Thr | Ser | Trp | Thr | Gln | Ile | Asp | Ala | Ile | Leu |      |
|    | 770 |     |     |     |     | 775 |     |     |     |     |     | 780 |     |     |     |     |      |
| 60 | CAA | TGG | TTA | CAG | ATG | TCT | TCG | GCC | TTG | GCG | GTT | TCT | CCA | CTG | GAT | CTG | 2400 |
|    | Gln | Trp | Leu | Gln | Met | Ser | Ser | Ala | Leu | Ala | Val | Ser | Pro | Leu | Asp | Leu |      |
|    | 785 |     |     |     |     | 790 |     |     |     |     | 795 |     |     |     |     | 800 |      |
| 65 | GCA | GGG | ATG | ATG | GCC | CTG | AAA | TAT | GGG | ATA | GAT | CAT | AAC | TAT | GCT | GCC | 2448 |
|    | Ala | Gly | Met | Met | Ala | Leu | Lys | Tyr | Gly | Ile | Asp | His | Asn | Tyr | Ala | Ala |      |
|    |     |     |     |     | 805 |     |     |     |     | 810 |     |     |     |     | 815 |     |      |
| 70 | TGG | CAA | GCT | GCG | GCG | GCT | GCG | CTG | ATG | GCT | GAT | CAT | GCT | AAT | CAG | GCA | 2496 |
|    | Trp | Gln | Ala | Ala | Ala | Ala | Ala | Leu | Met | Ala | Asp | His | Ala | Asn | Gln | Ala |      |
|    |     |     |     | 820 |     |     |     |     | 825 |     |     |     |     | 830 |     |     |      |
| 75 | CAG | AAA | AAA | CTG | GAT | GAG | ACG | TTC | AGT | AAG | GCA | TTA | TGT | AAC | TAT | TAT | 2544 |
|    | Gln | Lys | Lys | Leu | Asp | Glu | Thr | Phe | Ser | Lys | Ala | Leu | Cys | Asn | Tyr | Tyr |      |
|    |     |     | 835 |     |     |     |     | 840 |     |     |     |     | 845 |     |     |     |      |
| 80 | ATT | AAT | GCT | GTT | GTC | GAT | AGT | GCT | GCT | GGA | GTA | CGT | GAT | CGT | AAC | GGT | 2592 |
|    | Ile | Asn | Ala | Val | Val | Asp | Ser | Ala | Ala | Gly | Val | Arg | Asp | Arg | Asn | Gly |      |
|    | 850 |     |     |     |     | 855 |     |     |     |     |     | 860 |     |     |     |     |      |
| 85 | TTA | TAT | ACC | TAT | TTG | CTG | ATT | GAT | AAT | CAG | GTT | TCT | GCC | GAT | GTG | ATC | 2640 |
|    | Leu | Tyr | Thr | Tyr | Leu | Leu | Ile | Asp | Asn | Gln | Val | Ser | Ala | Asp | Val | Ile |      |

|    |      |      |      |      |      |      |      |      |      |      |      |      |     |      |      |     |      |
|----|------|------|------|------|------|------|------|------|------|------|------|------|-----|------|------|-----|------|
|    | 265  |      |      |      | 370  |      |      |      |      | 375  |      |      |     |      | 380  |     |      |
|    | ACT  | TCA  | CGT  | ATT  | GCA  | GAA  | GCT  | ATC  | GCC  | GGT  | ATT  | CAA  | CTG | TAC  | GTT  | AAC | 2688 |
| 5  | Thr  | Ser  | Arg  | Ile  | Ala  | Glu  | Ala  | Ile  | Ala  | Gly  | Ile  | Gln  | Leu | Tyr  | Val  | Asn |      |
|    |      |      |      |      | 885  |      |      |      |      | 890  |      |      |     |      | 895  |     |      |
|    | CGG  | GCT  | TTA  | AAC  | CGA  | GAT  | GAA  | GGT  | CAG  | CTT  | GCA  | TCG  | GAC | GTT  | AGT  | ACC | 2736 |
|    | Arg  | Ala  | Leu  | Asn  | Arg  | Asp  | Glu  | Gly  | Gln  | Leu  | Ala  | Ser  | Asp | Val  | Ser  | Thr |      |
| 10 |      |      |      | 900  |      |      |      |      | 905  |      |      |      |     | 910  |      |     |      |
|    | CGT  | CAG  | TTC  | TTC  | ACT  | GAC  | TGG  | GAA  | CGT  | TAC  | AAT  | AAA  | CGT | TAC  | AGT  | ACT | 2784 |
|    | Arg  | Gln  | Phe  | Phe  | Thr  | Asp  | Trp  | Glu  | Arg  | Tyr  | Asn  | Lys  | Arg | Tyr  | Ser  | Thr |      |
|    |      |      | 915  |      |      |      |      | 920  |      |      |      |      | 925 |      |      |     |      |
| 15 | TGG  | GCT  | GGT  | GTC  | TCT  | GAA  | CTG  | GTC  | TAT  | TAT  | CCA  | GAA  | AAC | TAT  | GTT  | GAT | 2832 |
|    | Trp  | Ala  | Gly  | Val  | Ser  | Glu  | Leu  | Val  | Tyr  | Tyr  | Pro  | Glu  | Asn | Tyr  | Val  | Asp |      |
|    |      | 930  |      |      |      |      | 935  |      |      |      |      | 940  |     |      |      |     |      |
| 20 | CCC  | ACT  | CAG  | CGC  | ATT  | GGG  | CAA  | ACC  | AAA  | ATG  | ATG  | GAT  | GCG | CTG  | TTG  | CAA | 2880 |
|    | Pro  | Thr  | Gln  | Arg  | Ile  | Gly  | Gln  | Thr  | Lys  | Met  | Met  | Asp  | Ala | Leu  | Leu  | Gln |      |
|    | 945  |      |      |      |      | 950  |      |      |      |      | 955  |      |     |      | 960  |     |      |
|    | TCC  | ATC  | AAC  | CAG  | AGC  | CAG  | CTA  | AAT  | GCG  | GAT  | ACG  | GTG  | GAA | GAT  | GCT  | TTC | 2928 |
| 25 | Ser  | Ile  | Asn  | Gln  | Ser  | Gln  | Leu  | Asn  | Ala  | Asp  | Thr  | Val  | Glu | Asp  | Ala  | Phe |      |
|    |      |      |      |      | 965  |      |      |      |      | 970  |      |      |     |      | 975  |     |      |
|    | AAA  | ACT  | TAT  | TTG  | ACC  | AGC  | TTT  | GAG  | CAG  | GTA  | GCA  | AAT  | CTG | AAA  | GTA  | ATT | 2976 |
|    | Lys  | Thr  | Tyr  | Leu  | Thr  | Ser  | Phe  | Glu  | Gln  | Val  | Ala  | Asn  | Leu | Lys  | Val  | Ile |      |
| 30 |      |      |      | 980  |      |      |      | 985  |      |      |      |      | 990 |      |      |     |      |
|    | AGT  | GCT  | TAC  | CAC  | GAT  | AAT  | GTG  | AAT  | GTG  | GAT  | CAA  | GGA  | TTA | ACT  | TAT  | TTT | 3024 |
|    | Ser  | Ala  | Tyr  | His  | Asp  | Asn  | Val  | Asn  | Val  | Asp  | Gln  | Gly  | Leu | Thr  | Tyr  | Phe |      |
|    |      |      | 995  |      |      |      | 1000 |      |      |      |      | 1005 |     |      |      |     |      |
| 35 | ATC  | GGT  | ATC  | GAC  | CAA  | GCA  | GCT  | CCG  | GGT  | ACG  | TAT  | TAC  | TGG | CGT  | AGT  | GTT | 3072 |
|    | Ile  | Gly  | Ile  | Asp  | Gln  | Ala  | Ala  | Pro  | Gly  | Thr  | Tyr  | Tyr  | Trp | Arg  | Ser  | Val |      |
|    |      | 1010 |      |      |      |      | 1015 |      |      |      |      | 1020 |     |      |      |     |      |
| 40 | GAT  | CAC  | AGC  | AAA  | TGT  | GAA  | AAT  | GGC  | AAG  | TTT  | GCC  | GCT  | AAT | GCT  | TGG  | GGT | 3120 |
|    | Asp  | His  | Ser  | Lys  | Cys  | Glu  | Asn  | Gly  | Lys  | Phe  | Ala  | Ala  | Asn | Ala  | Trp  | Gly |      |
|    | 1025 |      |      |      |      | 1030 |      |      |      | 1035 |      |      |     |      | 1040 |     |      |
|    | GAG  | TGG  | AAT  | AAA  | ATT  | ACC  | TGT  | GCT  | GTC  | AAT  | CCT  | TGG  | AAA | AAT  | ATC  | ATC | 3168 |
| 45 | Glu  | Trp  | Asn  | Lys  | Ile  | Thr  | Cys  | Ala  | Val  | Asn  | Pro  | Trp  | Lys | Asn  | Ile  | Ile |      |
|    |      |      |      |      | 1045 |      |      |      | 1050 |      |      |      |     |      | 1055 |     |      |
|    | CGT  | CCG  | GTT  | GTT  | TAT  | ATG  | TCC  | CGC  | TTA  | TAT  | CTG  | CTA  | TGG | CTG  | GAG  | CAG | 3216 |
|    | Arg  | Pro  | Val  | Val  | Tyr  | Met  | Ser  | Arg  | Leu  | Tyr  | Leu  | Leu  | Trp | Leu  | Glu  | Gln |      |
| 50 |      |      |      | 1060 |      |      |      |      | 1065 |      |      |      |     | 1070 |      |     |      |
|    | CAA  | TCA  | AAG  | AAA  | AGT  | GAT  | GAT  | GGT  | AAA  | ACC  | ACG  | ATT  | TAT | CAA  | TAT  | AAC | 3264 |
|    | Gln  | Ser  | Lys  | Lys  | Ser  | Asp  | Asp  | Gly  | Lys  | Thr  | Thr  | Ile  | Tyr | Gln  | Tyr  | Asn |      |
|    |      |      | 1075 |      |      |      |      | 1080 |      |      |      | 1085 |     |      |      |     |      |
| 55 | TTA  | AAA  | CTG  | GCT  | CAT  | ATT  | CGT  | TAC  | GAC  | GGT  | AGT  | TGG  | AAT | ACA  | CCA  | TTT | 3312 |
|    | Leu  | Lys  | Leu  | Ala  | His  | Ile  | Arg  | Tyr  | Asp  | Gly  | Ser  | Trp  | Asn | Thr  | Pro  | Phe |      |
|    |      | 1090 |      |      |      |      | 1095 |      |      |      |      | 1100 |     |      |      |     |      |
| 60 | ACT  | TTT  | GAT  | GTG  | ACA  | GAA  | AAG  | GTA  | AAA  | AAT  | TAC  | ACG  | TCG | AGT  | ACT  | GAT | 3360 |
|    | Thr  | Phe  | Asp  | Val  | Thr  | Glu  | Lys  | Val  | Lys  | Asn  | Tyr  | Thr  | Ser | Ser  | Thr  | Asp |      |
|    | 1105 |      |      |      |      | 1110 |      |      |      |      | 1115 |      |     |      | 1120 |     |      |
|    | GCT  | GCT  | GAA  | TCT  | TTA  | GGG  | TTG  | TAT  | TGT  | ACT  | GGT  | TAT  | CAA | GGG  | GAA  | GAC | 3408 |
| 65 | Ala  | Ala  | Glu  | Ser  | Leu  | Gly  | Leu  | Tyr  | Cys  | Thr  | Gly  | Tyr  | Gln | Gly  | Glu  | Asp |      |
|    |      |      |      |      | 1125 |      |      |      |      | 1130 |      |      |     |      | 1135 |     |      |
|    | ACT  | CTA  | TTA  | GTT  | ATG  | TTC  | TAT  | TCG  | ATG  | CAG  | AGT  | AGT  | TAT | AGC  | TCC  | TAT | 3456 |
|    | Thr  | Leu  | Leu  | Val  | Met  | Phe  | Tyr  | Ser  | Met  | Gln  | Ser  | Ser  | Tyr | Ser  | Ser  | Tyr |      |
| 70 |      |      |      | 1140 |      |      |      |      | 1145 |      |      |      |     | 1150 |      |     |      |
|    | ACC  | GAT  | AAT  | AAT  | GCG  | CCG  | GTC  | ACT  | GGG  | CTA  | TAT  | ATT  | TTC | GCT  | GAT  | ATG | 3504 |

Thr Asp Asn Asn Ala Pro Val Thr Gly Leu Tyr Ile Phe Ala Asp Met  
 1155 1160 1165  
 5 TCA TCA GAC AAT ATG ACG AAT GCA CAA GCA ACT AAC TAT TGG AAT AAC 3550  
 Ser Ser Asp Asn Met Thr Asn Ala Gln Ala Thr Asn Tyr Trp Asn Asn  
 1170 1175 1180  
 10 AGT TAT CCG CAA TTT GAT ACT GTG ATG GCA GAT CCG GAT AGC GAC AAT 3600  
 Ser Tyr Pro Gln Phe Asp Thr Val Met Ala Asp Pro Asp Ser Asp Asn  
 1185 1190 1195 1200  
 15 AAA AAA GTC ATA ACC AGA AGA GTT AAT AAC CGT TAT GCG GAG GAT TAT 3648  
 Lys Lys Val Ile Thr Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr  
 1205 1210 1215  
 GAA ATT CCT TCC TCT GTG ACA AGT AAC AGT AAT TAT TCT TGG GGT GAT 3696  
 Glu Ile Pro Ser Ser Val Thr Ser Asn Ser Asn Tyr Ser Trp Gly Asp  
 1220 1225 1230  
 20 CAC AGT TTA ACC ATG CTT TAT GGT GGT AGT GTT CCT AAT ATT ACT TTT 3744  
 His Ser Leu Thr Met Leu Tyr Gly Gly Ser Val Pro Asn Ile Thr Phe  
 1235 1240 1245  
 25 GAA TCG GCG GCA GAA GAT TTA AGG CTA TCT ACC AAT ATG GCA TTG AGT 3792  
 Glu Ser Ala Ala Glu Asp Leu Arg Leu Ser Thr Asn Met Ala Leu Ser  
 1250 1255 1260  
 30 ATT ATT CAT AAT GGA TAT GCG GGA ACC CGC CGT ATA CAA TGT AAT CTT 3840  
 Ile Ile His Asn Gly Tyr Ala Gly Thr Arg Arg Ile Gln Cys Asn Leu  
 1265 1270 1275 1280  
 35 ATG AAA CAA TAC GCT TCA TTA GGT GAT AAA TTT ATA ATT TAT GAT TCA 3888  
 Met Lys Gln Tyr Ala Ser Leu Gly Asp Lys Phe Ile Ile Tyr Asp Ser  
 1285 1290 1295  
 TCA TTT GAT GAT GCA AAC CGT TTT AAT CTG GTG CCA TTG TTT AAA TTC 3936  
 Ser Phe Asp Asp Ala Asn Arg Phe Asn Leu Val Pro Leu Phe Lys Phe  
 1300 1305 1310  
 40 GGA AAA GAC GAG AAC TCA GAT GAT AGT ATT TGT ATA TAT AAT GAA AAC 3984  
 Gly Lys Asp Glu Asn Ser Asp Asp Ser Ile Cys Ile Tyr Asn Glu Asn  
 1315 1320 1325  
 45 CCT TCC TCT GAA GAT AAG AAG TGG TAT TTT TCT TCG AAA GAT GAC AAT 4032  
 Pro Ser Ser Glu Asp Lys Lys Trp Tyr Phe Ser Ser Lys Asp Asp Asn  
 1330 1335 1340  
 50 AAA ACA GCG GAT TAT AAT GGT GGA ACT CAA TGT ATA GAT GCT GGA ACC 4080  
 Lys Thr Ala Asp Tyr Asn Gly Gly Thr Gln Cys Ile Asp Ala Gly Thr  
 1345 1350 1355 1360  
 AGT AAC AAA GAT TTT TAT TAT AAT CTC CAG GAG ATT GAA GTA ATT AGT 4128  
 Ser Asn Lys Asp Phe Tyr Asn Leu Gln Glu Ile Glu Val Ile Ser  
 1365 1370 1375  
 55 GTT ACT GGT GGG TAT TGG TCG AGT TAT AAA ATA TCC AAC CCG ATT AAT 4176  
 Val Thr Gly Gly Tyr Trp Ser Ser Tyr Lys Ile Ser Asn Pro Ile Asn  
 1380 1385 1390  
 60 ATC AAT ACG GGC ATT GAT AGT GCT AAA GTA AAA GTC ACC GTA AAA GCG 4224  
 Ile Asn Thr Gly Ile Asp Ser Ala Lys Val Lys Val Thr Val Lys Ala  
 1395 1400 1405  
 65 GGT GGT GAC GAT CAA ATC TTT ACT GCT GAT AAT AGT ACC TAT GTT CCT 4272  
 Gly Gly Asp Asp Gln Ile Phe Thr Ala Asp Asn Ser Thr Tyr Val Pro  
 1410 1415 1420  
 70 CAG CAA CCG GCA CCC AGT TTT GAG GAG ATG ATT TAT CAG TTC AAT AAC 4320  
 Gln Gln Pro Ala Pro Ser Phe Glu Glu Met Ile Tyr Gln Phe Asn Asn  
 1425 1430 1435 1440

|    |     |      |      |      |      |      |      |      |      |      |     |      |      |      |      |      |      |
|----|-----|------|------|------|------|------|------|------|------|------|-----|------|------|------|------|------|------|
|    | CTG | ACA  | ATA  | GAT  | TGT  | AAG  | AAT  | TTA  | AAT  | TTC  | ATC | GAC  | AAT  | CAG  | GCA  | CAT  | 4163 |
|    | Leu | Thr  | Ile  | Asp  | Cys  | Lys  | Asn  | Leu  | Asn  | Phe  | Ile | Asp  | Asn  | Gln  | Ala  | His  |      |
|    |     |      |      |      | 1445 |      |      |      |      | 1450 |     |      |      |      | 1455 |      |      |
| 5  | ATT | GAG  | ATT  | GAT  | TTC  | ACC  | GCT  | ACG  | GCA  | CAA  | GAT | GGC  | CGA  | TTC  | TTG  | GGT  | 4416 |
|    | Ile | Glu  | Ile  | Asp  | Phe  | Thr  | Ala  | Thr  | Ala  | Gln  | Asp | Gly  | Arg  | Phe  | Leu  | Gly  |      |
|    |     |      |      |      | 1460 |      |      |      |      | 1465 |     |      |      |      | 1470 |      |      |
| 10 | GCA | GAA  | ACT  | TTT  | ATT  | ATC  | CCG  | GTA  | ACT  | AAA  | AAA | GTT  | CTC  | GGT  | ACT  | GAG  | 4464 |
|    | Ala | Glu  | Thr  | Phe  | Ile  | Ile  | Pro  | Val  | Thr  | Lys  | Lys | Val  | Leu  | Gly  | Thr  | Glu  |      |
|    |     |      |      |      | 1475 |      |      |      |      | 1480 |     |      |      |      | 1485 |      |      |
| 15 | AAC | GTG  | ATT  | GCG  | TTA  | TAT  | AGC  | GAA  | AAT  | AAC  | GGT | GTT  | CAA  | TAT  | ATG  | CAA  | 4512 |
|    | Asn | Val  | Ile  | Ala  | Leu  | Tyr  | Ser  | Glu  | Asn  | Asn  | Gly | Val  | Gln  | Tyr  | Met  | Gln  |      |
|    |     | 1490 |      |      |      |      |      | 1495 |      |      |     |      | 1500 |      |      |      |      |
| 20 | ATT | GGC  | GCA  | TAT  | CGT  | ACC  | CGT  | TTG  | AAT  | ACG  | TTA | TTC  | GCT  | CAA  | CAG  | TTG  | 4560 |
|    | Ile | Gly  | Ala  | Tyr  | Arg  | Thr  | Arg  | Leu  | Asn  | Thr  | Leu | Phe  | Ala  | Gln  | Gln  | Leu  |      |
|    |     | 1505 |      |      |      |      | 1510 |      |      |      |     | 1515 |      |      |      | 1520 |      |
| 25 | GTT | AGC  | CGT  | GCT  | AAT  | CGT  | GGC  | ATT  | GAT  | GCA  | GTG | CTC  | AGT  | ATG  | GAA  | ACT  | 4608 |
|    | Val | Ser  | Arg  | Ala  | Asn  | Arg  | Gly  | Ile  | Asp  | Ala  | Val | Leu  | Ser  | Met  | Glu  | Thr  |      |
|    |     |      |      |      |      | 1525 |      |      |      | 1530 |     |      |      |      | 1535 |      |      |
| 30 | CAG | AAT  | ATT  | CAG  | GAA  | CCG  | CAA  | TTA  | GGA  | GCG  | GGC | ACA  | TAT  | GTG  | CAG  | CTT  | 4656 |
|    | Gln | Asn  | Ile  | Gln  | Glu  | Pro  | Gln  | Leu  | Gly  | Ala  | Gly | Thr  | Tyr  | Val  | Gln  | Leu  |      |
|    |     |      |      |      | 1540 |      |      |      | 1545 |      |     |      |      |      | 1550 |      |      |
| 35 | GTG | TTG  | GAT  | AAA  | TAT  | GAT  | GAG  | TCT  | ATT  | CAT  | GGC | ACT  | AAT  | AAA  | AGC  | TTT  | 4704 |
|    | Val | Leu  | Asp  | Lys  | Tyr  | Asp  | Glu  | Ser  | Ile  | His  | Gly | Thr  | Asn  | Lys  | Ser  | Phe  |      |
|    |     |      | 1555 |      |      |      |      | 1560 |      |      |     |      |      | 1565 |      |      |      |
| 40 | GCT | ATT  | GAA  | TAT  | GTT  | GAT  | ATA  | TTT  | AAA  | GAG  | AAC | GAT  | AGT  | TTT  | GTG  | ATT  | 4752 |
|    | Ala | Ile  | Glu  | Tyr  | Val  | Asp  | Ile  | Phe  | Lys  | Glu  | Asn | Asp  | Ser  | Phe  | Val  | Ile  |      |
|    |     | 1570 |      |      |      |      | 1575 |      |      |      |     |      | 1580 |      |      |      |      |
| 45 | TAT | CAA  | GGA  | GAA  | CTT  | AGC  | GAA  | ACA  | AGT  | CAA  | ACT | GTT  | GTG  | AAA  | GTT  | TTC  | 4800 |
|    | Tyr | Gln  | Gly  | Glu  | Leu  | Ser  | Glu  | Thr  | Ser  | Gln  | Thr | Val  | Val  | Lys  | Val  | Phe  |      |
|    |     | 1585 |      |      |      |      | 1590 |      |      |      |     | 1595 |      |      |      | 1600 |      |
| 50 | TTA | TCC  | TAT  | TTT  | ATA  | GAG  | GCG  | ACT  | GGA  | AAT  | AAG | AAC  | CAC  | TTA  | TGG  | GTA  | 4848 |
|    | Leu | Ser  | Tyr  | Phe  | Ile  | Glu  | Ala  | Thr  | Gly  | Asn  | Lys | Asn  | His  | Leu  | Trp  | Val  |      |
|    |     |      |      |      | 1605 |      |      |      |      | 1610 |     |      |      |      | 1615 |      |      |
| 55 | CGT | GCT  | AAA  | TAC  | CAA  | AAG  | GAA  | ACG  | ACT  | GAT  | AAG | ATC  | TTG  | TTC  | GAC  | CGT  | 4896 |
|    | Arg | Ala  | Lys  | Tyr  | Gln  | Lys  | Glu  | Thr  | Thr  | Asp  | Lys | Ile  | Leu  | Phe  | Asp  | Arg  |      |
|    |     |      |      | 1620 |      |      |      |      |      | 1625 |     |      |      |      | 1630 |      |      |
| 60 | ACT | GAT  | GAG  | AAA  | GAT  | CCG  | CAC  | GGT  | TGG  | TTT  | CTC | AGC  | GAC  | GAT  | CAC  | AAG  | 4944 |
|    | Thr | Asp  | Glu  | Lys  | Asp  | Pro  | His  | Gly  | Trp  | Phe  | Leu | Ser  | Asp  | Asp  | His  | Lys  |      |
|    |     |      | 1635 |      |      |      |      | 1640 |      |      |     |      |      | 1645 |      |      |      |
| 65 | ACC | TTT  | AGT  | GGT  | CTC  | TCT  | TCC  | GCA  | CAG  | GCA  | TTA | AAG  | AAC  | GAC  | AGT  | GAA  | 4992 |
|    | Thr | Phe  | Ser  | Gly  | Leu  | Ser  | Ser  | Ala  | Gln  | Ala  | Leu | Lys  | Asn  | Asp  | Ser  | Glu  |      |
|    |     | 1650 |      |      |      |      |      | 1655 |      |      |     |      | 1660 |      |      |      |      |
| 70 | CCG | ATG  | GAT  | TTC  | TCT  | GGC  | GCC  | AAT  | GCT  | CTC  | TAT | TTC  | TGG  | GAA  | CTG  | TTC  | 5040 |
|    | Pro | Met  | Asp  | Phe  | Ser  | Gly  | Ala  | Asn  | Ala  | Leu  | Tyr | Phe  | Trp  | Glu  | Leu  | Phe  |      |
|    |     | 1665 |      |      |      |      | 1670 |      |      |      |     | 1675 |      |      |      | 1680 |      |
| 75 | TAT | TAC  | ACG  | CCG  | ATG  | ATG  | ATG  | GCT  | CAT  | CGT  | TTG | TTG  | CAG  | GAA  | CAG  | AAT  | 5088 |
|    | Tyr | Tyr  | Thr  | Pro  | Met  | Met  | Met  | Ala  | His  | Arg  | Leu | Leu  | Gln  | Glu  | Gln  | Asn  |      |
|    |     |      |      |      | 1685 |      |      |      |      | 1690 |     |      |      |      | 1695 |      |      |
| 80 | TTT | GAT  | GCG  | GCG  | AAC  | CAT  | TGG  | TTC  | CGT  | TAT  | GTC | TGG  | AGT  | CCA  | TCC  | GGT  | 5136 |
|    | Phe | Asp  | Ala  | Ala  | Asn  | His  | Trp  | Phe  | Arg  | Tyr  | Val | Trp  | Ser  | Pro  | Ser  | Gly  |      |
|    |     |      |      |      | 1700 |      |      |      | 1705 |      |     |      |      |      | 1710 |      |      |
| 85 | TAT | ATC  | GTT  | GAT  | GGT  | AAA  | ATT  | GCT  | ATC  | TAC  | CAC | TGG  | AAC  | GTG  | CGA  | CCG  | 5184 |
|    | Tyr | Ile  | Val  | Asp  | Gly  | Lys  | Ile  | Ala  | Ile  | Tyr  | His | Trp  | Asn  | Val  | Arg  | Pro  |      |
|    |     |      | 1715 |      |      |      |      | 1720 |      |      |     |      |      | 1725 |      |      |      |

CTG GAA GAA GAC ACC AGT TGG AAT GCA CAA CAA CTG GAC TCC ACC GAT 5131  
 Leu Glu Glu Asp Thr Ser Trp Asn Ala Gln Gln Leu Asp Ser Thr Asp  
 1730 1735 1740

5 CCA GAT GCT GTA GCC CAA GAT GAT CCG ATG CAC TAC AAG GTG GCT ACC 5280  
 Pro Asp Ala Val Ala Gln Asp Asp Pro Met His Tyr Lys Val Ala Thr  
 1745 1750 1755 1760

10 TTT ATG GCG ACG TTG GAT CTG CTA ATG GCC CGT GGT GAT GCT GCT TAC 5328  
 Phe Met Ala Thr Leu Asp Leu Leu Met Ala Arg Gly Asp Ala Ala Tyr  
 1765 1770 1775

15 CGC CAG TTA GAG CGT GAT ACG TTG GCT GAA GCT AAA ATG TGG TAT ACA 5376  
 Arg Gln Leu Gln Arg Asp Thr Leu Ala Glu Ala Lys Met Trp Tyr Thr  
 1780 1785 1790

20 CAG GCG CTT AAT CTG TTG GGT GAT GAG CCA CAA GTG ATG CTG AGT ACG 5424  
 Gln Ala Leu Asn Leu Leu Gly Asp Glu Pro Gln Val Met Leu Ser Thr  
 1795 1800 1805

ACT TGG GCT AAT CCA ACA TTG GGT AAT GCT GCT TCA AAA ACC ACA CAG 5472  
 Thr Trp Ala Asn Pro Thr Leu Gly Asn Ala Ala Ser Lys Thr Thr Gln  
 1810 1815 1820

25 CAG GTT CGT CAG CAA GTG CTT ACC CAG TTG CGT CTC AAT AGC AGG GTA 5520  
 Gln Val Arg Gln Gln Val Leu Thr Gln Leu Arg Leu Asn Ser Arg Val  
 1825 1830 1835 1840

30 AAA ACC CCG TTG 5532  
 Lys Thr Pro Leu  
 1844

35 (2) INFORMATION FOR SEQ ID NO:53:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1844 amino acids  
 (B) TYPE: amino acids  
 (C) STRANDEDNESS: single  
 40 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53 (TcbA<sub>ii</sub>):

| Features    | From | To   | Description                |
|-------------|------|------|----------------------------|
| Peptide     | 1    | 1844 | TcbA <sub>ii</sub> peptide |
| Fragment    | 1    | 11   | (SEQ ID NO:1)              |
| Fragment    | 978  | 990  | (SEQ ID NO:23)             |
| 50 Fragment | 1387 | 1401 | (SEQ ID NO:22)             |
| Fragment    | 1484 | 1505 | (SEQ ID NO:24)             |
| Fragment    | 1527 | 1552 | (SEQ ID NO:21)             |

55 Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala Asp Asn Tyr  
 1 5 10 15

Ala Ala Pro Gly Ser Val Ala Ser Met Phe Ser Pro Ala Ala Tyr Leu  
 20 25 30

60 Thr Glu Leu Tyr Arg Glu Ala Lys Asn Leu His Asp Ser Ser Ser Ile  
 35 40 45

Tyr Tyr Leu Asp Lys Arg Arg Pro Asp Leu Ala Ser Leu Met Leu Ser  
 50 55 60

65 Gln Lys Asn Met Asp Glu Glu Ile Ser Thr Leu Ala Leu Ser Asn Glu  
 65 70 75 80

Leu Cys Leu Ala Gly Ile Glu Thr Lys Thr Gly Lys Ser Gln Asp Glu

|    |  |  |  |  |                                                                 |  |  |  |  |     |  |  |  |  |  |     |
|----|--|--|--|--|-----------------------------------------------------------------|--|--|--|--|-----|--|--|--|--|--|-----|
|    |  |  |  |  | 85                                                              |  |  |  |  | 90  |  |  |  |  |  | 95  |
|    |  |  |  |  | Val Met Asp Met Leu Ser Thr Tyr Arg Leu Ser Gly Glu Thr Pro Tyr |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 100                                                             |  |  |  |  | 105 |  |  |  |  |  | 110 |
| 5  |  |  |  |  | His His Ala Tyr Glu Thr Val Arg Glu Ile Val His Glu Arg Asp Pro |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 115                                                             |  |  |  |  | 120 |  |  |  |  |  | 125 |
|    |  |  |  |  | Gly Phe Arg His Leu Ser Gln Ala Pro Ile Val Ala Ala Lys Leu Asp |  |  |  |  |     |  |  |  |  |  |     |
| 10 |  |  |  |  | 130                                                             |  |  |  |  | 135 |  |  |  |  |  | 140 |
|    |  |  |  |  | Pro Val Thr Leu Leu Gly Ile Ser Ser His Ile Ser Pro Glu Leu Tyr |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 145                                                             |  |  |  |  | 150 |  |  |  |  |  | 155 |
|    |  |  |  |  | Asn Leu Leu Ile Glu Glu Ile Pro Glu Lys Asp Glu Ala Ala Leu Asp |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 165                                                             |  |  |  |  | 170 |  |  |  |  |  | 175 |
|    |  |  |  |  | Thr Leu Tyr Lys Thr Asn Phe Gly Asp Ile Thr Thr Ala Gln Leu Met |  |  |  |  | 185 |  |  |  |  |  |     |
| 20 |  |  |  |  | 180                                                             |  |  |  |  |     |  |  |  |  |  | 190 |
|    |  |  |  |  | Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr Gly Val Ser Pro Glu Asp Ile |  |  |  |  | 200 |  |  |  |  |  |     |
|    |  |  |  |  | 195                                                             |  |  |  |  |     |  |  |  |  |  | 205 |
|    |  |  |  |  | Ala Tyr Val Thr Thr Ser Leu Ser His Val Gly Tyr Ser Ser Asp Ile |  |  |  |  |     |  |  |  |  |  |     |
| 25 |  |  |  |  | 210                                                             |  |  |  |  | 215 |  |  |  |  |  | 220 |
|    |  |  |  |  | Leu Val Ile Pro Leu Val Asp Gly Val Gly Lys Met Glu Val Val Arg |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 225                                                             |  |  |  |  | 230 |  |  |  |  |  | 235 |
|    |  |  |  |  | Val Thr Arg Thr Pro Ser Asp Asn Tyr Thr Ser Gln Thr Asn Tyr Ile |  |  |  |  |     |  |  |  |  |  |     |
| 30 |  |  |  |  | 245                                                             |  |  |  |  | 250 |  |  |  |  |  | 255 |
|    |  |  |  |  | Glu Leu Tyr Pro Gln Gly Gly Asp Asn Tyr Leu Ile Lys Tyr Asn Leu |  |  |  |  | 265 |  |  |  |  |  |     |
|    |  |  |  |  | 260                                                             |  |  |  |  |     |  |  |  |  |  | 270 |
| 35 |  |  |  |  | Ser Asn Ser Phe Gly Leu Asp Asp Phe Tyr Leu Gln Tyr Lys Asp Gly |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 275                                                             |  |  |  |  | 280 |  |  |  |  |  | 285 |
|    |  |  |  |  | Ser Ala Asp Trp Thr Glu Ile Ala His Asn Pro Tyr Pro Asp Met Val |  |  |  |  |     |  |  |  |  |  |     |
| 40 |  |  |  |  | 290                                                             |  |  |  |  | 295 |  |  |  |  |  | 300 |
|    |  |  |  |  | Ile Asn Gln Lys Tyr Glu Ser Gln Ala Thr Ile Lys Arg Ser Asp Ser |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 305                                                             |  |  |  |  | 310 |  |  |  |  |  | 315 |
|    |  |  |  |  | Asp Asn Ile Leu Ser Ile Gly Leu Gln Arg Trp His Ser Gly Ser Tyr |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 325                                                             |  |  |  |  | 330 |  |  |  |  |  | 335 |
|    |  |  |  |  | Asn Phe Ala Ala Ala Asn Phe Lys Ile Asp Gln Tyr Ser Pro Lys Ala |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 340                                                             |  |  |  |  | 345 |  |  |  |  |  | 350 |
| 50 |  |  |  |  | Phe Leu Leu Lys Met Asn Lys Ala Ile Arg Leu Leu Lys Ala Thr Gly |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 355                                                             |  |  |  |  | 360 |  |  |  |  |  | 365 |
|    |  |  |  |  | Leu Ser Phe Ala Thr Leu Glu Arg Ile Val Asp Ser Val Asn Ser Thr |  |  |  |  |     |  |  |  |  |  |     |
| 55 |  |  |  |  | 370                                                             |  |  |  |  | 375 |  |  |  |  |  | 380 |
|    |  |  |  |  | Lys Ser Ile Thr Val Glu Val Leu Asn Lys Val Tyr Arg Val Lys Phe |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 385                                                             |  |  |  |  | 390 |  |  |  |  |  | 395 |
|    |  |  |  |  | Tyr Ile Asp Arg Tyr Gly Ile Ser Glu Glu Thr Ala Ala Ile Leu Ala |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 405                                                             |  |  |  |  | 410 |  |  |  |  |  | 415 |
|    |  |  |  |  | Asn Ile Asn Ile Ser Gln Gln Ala Val Gly Asn Gln Leu Ser Gln Phe |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 420                                                             |  |  |  |  | 425 |  |  |  |  |  | 430 |
| 65 |  |  |  |  | Glu Gln Leu Phe Asn His Pro Pro Leu Asn Gly Ile Arg Tyr Glu Ile |  |  |  |  |     |  |  |  |  |  |     |
|    |  |  |  |  | 435                                                             |  |  |  |  | 440 |  |  |  |  |  | 445 |
|    |  |  |  |  | Ser Glu Asp Asn Ser Lys His Leu Pro Asn Pro Asp Leu Asn Leu Lys |  |  |  |  |     |  |  |  |  |  |     |
| 70 |  |  |  |  | 450                                                             |  |  |  |  | 455 |  |  |  |  |  | 460 |



|    |                                                                 |     |     |     |     |
|----|-----------------------------------------------------------------|-----|-----|-----|-----|
|    | Pro Asp Ser Thr Gly Asp Asp Gln Arg Lys Ala Val Leu Lys Arg Ala | 465 | 470 | 475 | 480 |
| 5  | Phe Gln Val Asn Ala Ser Glu Leu Tyr Gln Met Leu Leu Ile Thr Asp | 485 | 490 | 495 |     |
|    | Arg Lys Glu Asp Gly Val Ile Lys Asn Asn Leu Glu Asn Leu Ser Asp | 500 | 505 | 510 |     |
| 10 | Leu Tyr Leu Val Ser Leu Leu Ala Gln Ile His Asn Leu Thr Ile Ala | 515 | 520 | 525 |     |
|    | Glu Leu Asn Ile Leu Leu Val Ile Cys Gly Tyr Gly Asp Thr Asn Ile | 530 | 535 | 540 |     |
| 15 | Tyr Gln Ile Thr Asp Asp Asn Leu Ala Lys Ile Val Glu Thr Leu Leu | 545 | 550 | 555 | 560 |
|    | Trp Ile Thr Gln Trp Leu Lys Thr Gln Lys Trp Thr Val Thr Asp Leu | 565 | 570 | 575 |     |
| 20 | Phe Leu Met Thr Thr Ala Thr Tyr Ser Thr Thr Leu Thr Pro Glu Ile | 580 | 585 | 590 |     |
|    | Ser Asn Leu Thr Ala Thr Leu Ser Ser Thr Leu His Gly Lys Glu Ser | 595 | 600 | 605 |     |
|    | Leu Ile Gly Glu Asp Leu Lys Arg Ala Met Ala Pro Cys Phe Thr Ser | 610 | 615 | 620 |     |
| 30 | Ala Leu His Leu Thr Ser Gln Glu Val Ala Tyr Asp Leu Leu Leu Trp | 625 | 630 | 635 | 640 |
|    | Ile Asp Gln Ile Gln Pro Ala Gln Ile Thr Val Asp Gly Phe Trp Glu | 645 | 650 | 655 |     |
| 35 | Glu Val Gln Thr Thr Pro Thr Ser Leu Lys Val Ile Thr Phe Ala Gln | 660 | 665 | 670 |     |
|    | Val Leu Ala Gln Leu Ser Leu Ile Tyr Arg Arg Ile Gly Leu Ser Glu | 675 | 680 | 685 |     |
|    | Thr Glu Leu Ser Leu Ile Val Thr Gln Ser Ser Leu Leu Val Ala Gly | 690 | 695 | 700 |     |
| 45 | Lys Ser Ile Leu Asp His Gly Leu Leu Thr Leu Met Ala Leu Glu Gly | 705 | 710 | 715 | 720 |
|    | Phe His Thr Trp Val Asn Gly Leu Gly Gln His Ala Ser Leu Ile Leu | 725 | 730 | 735 |     |
| 50 | Ala Ala Leu Lys Asp Gly Ala Leu Thr Val Thr Asp Val Ala Gln Ala | 740 | 745 | 750 |     |
|    | Met Asn Lys Glu Glu Ser Leu Leu Gln Met Ala Ala Asn Gln Val Glu | 755 | 760 | 765 |     |
|    | Lys Asp Leu Thr Lys Leu Thr Ser Trp Thr Gln Ile Asp Ala Ile Leu | 770 | 775 | 780 |     |
| 60 | Gln Trp Leu Gln Met Ser Ser Ala Leu Ala Val Ser Pro Leu Asp Leu | 785 | 790 | 795 | 800 |
|    | Ala Gly Met Met Ala Leu Lys Tyr Gly Ile Asp His Asn Tyr Ala Ala | 805 | 810 | 815 |     |
| 65 | Trp Gln Ala Ala Ala Ala Leu Met Ala Asp His Ala Asn Gln Ala     | 820 | 825 | 830 |     |
| 70 | Gln Lys Lys Leu Asp Glu Thr Phe Ser Lys Ala Leu Cys Asn Tyr Tyr | 835 | 840 | 845 |     |

Ile Asn Ala Val Val Asp Ser Ala Ala Gly Val Arg Asp Arg Asn Gly  
 350 855 860  
 5 Leu Tyr Thr Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Asp Val Ile  
 865 870 875 880  
 Thr Ser Arg Ile Ala Glu Ala Ile Ala Gly Ile Gln Leu Tyr Val Asn  
 885 890 895  
 10 Arg Ala Leu Asn Arg Asp Glu Gly Gln Leu Ala Ser Asp Val Ser Thr  
 900 905 910  
 15 Arg Gln Phe Phe Thr Asp Trp Glu Arg Tyr Asn Lys Arg Tyr Ser Thr  
 915 920 925  
 Trp Ala Gly Val Ser Glu Leu Val Tyr Tyr Pro Glu Asn Tyr Val Asp  
 930 935 940  
 20 Pro Thr Gln Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln  
 945 950 955 960  
 Ser Ile Asn Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe  
 965 970 975  
 25 Lys Thr Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile  
 980 985 990  
 30 Ser Ala Tyr His Asp Asn Val Asn Val Asp Gln Gly Leu Thr Tyr Phe  
 995 1000 1005  
 Ile Gly Ile Asp Gln Ala Ala Pro Gly Thr Tyr Tyr Trp Arg Ser Val  
 1010 1015 1020  
 35 Asp His Ser Lys Cys Glu Asn Gly Lys Phe Ala Ala Asn Ala Trp Gly  
 1025 1030 1035 1040  
 Glu Trp Asn Lys Ile Thr Cys Ala Val Asn Pro Trp Lys Asn Ile Ile  
 1045 1050 1055  
 40 Arg Pro Val Val Tyr Met Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln  
 1060 1065 1070  
 45 Gln Ser Lys Lys Ser Asp Asp Gly Lys Thr Thr Ile Tyr Gln Tyr Asn  
 1075 1080 1085  
 Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Ser Trp Asn Thr Pro Phe  
 1090 1095 1100  
 50 Thr Phe Asp Val Thr Glu Lys Val Lys Asn Tyr Thr Ser Ser Thr Asp  
 1105 1110 1115 1120  
 Ala Ala Glu Ser Leu Gly Leu Tyr Cys Thr Gly Tyr Gln Gly Glu Asp  
 1125 1130 1135  
 55 Thr Leu Leu Val Met Phe Tyr Ser Met Gln Ser Ser Tyr Ser Ser Tyr  
 1140 1145 1150  
 60 Thr Asp Asn Asn Ala Pro Val Thr Gly Leu Tyr Ile Phe Ala Asp Met  
 1155 1160 1165  
 Ser Ser Asp Asn Met Thr Asn Ala Gln Ala Thr Asn Tyr Trp Asn Asn  
 1170 1175 1180  
 65 Ser Tyr Pro Gln Phe Asp Thr Val Met Ala Asp Pro Asp Ser Asp Asn  
 1185 1190 1195 1200  
 70 Lys Lys Val Ile Thr Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr  
 1205 1210 1215  
 Glu Ile Pro Ser Ser Val Thr Ser Asn Ser Asn Tyr Ser Trp Gly Asp

|    | 1220                                                                                   | 1225 | 1230 |
|----|----------------------------------------------------------------------------------------|------|------|
|    | His Ser Leu Thr Met Leu Tyr Gly Gly Ser Val Pro Asn Ile Thr Phe<br>1235 1240 1245      |      |      |
| 5  | Glu Ser Ala Ala Glu Asp Leu Arg Leu Ser Thr Asn Met Ala Leu Ser<br>1250 1255 1260      |      |      |
| 10 | Ile Ile His Asn Gly Tyr Ala Gly Thr Arg Arg Ile Gln Cys Asn Leu<br>1265 1270 1275 1280 |      |      |
|    | Met Lys Gln Tyr Ala Ser Leu Gly Asp Lys Phe Ile Ile Tyr Asp Ser<br>1285 1290 1295      |      |      |
| 15 | Ser Phe Asp Asp Ala Asn Arg Phe Asn Leu Val Pro Leu Phe Lys Phe<br>1300 1305 1310      |      |      |
|    | Gly Lys Asp Glu Asn Ser Asp Asp Ser Ile Cys Ile Tyr Asn Glu Asn<br>1315 1320 1325      |      |      |
| 20 | Pro Ser Ser Glu Asp Lys Lys Trp Tyr Phe Ser Ser Lys Asp Asp Asn<br>1330 1335 1340      |      |      |
| 25 | Lys Thr Ala Asp Tyr Asn Gly Gly Thr Gln Cys Ile Asp Ala Gly Thr<br>1345 1350 1355 1360 |      |      |
|    | Ser Asn Lys Asp Phe Tyr Tyr Asn Leu Gln Glu Ile Glu Val Ile Ser<br>1365 1370 1375      |      |      |
| 30 | Val Thr Gly Gly Tyr Trp Ser Ser Tyr Lys Ile Ser Asn Pro Ile Asn<br>1380 1385 1390      |      |      |
|    | Ile Asn Thr Gly Ile Asp Ser Ala Lys Val Lys Val Thr Val Lys Ala<br>1395 1400 1405      |      |      |
| 35 | Gly Gly Asp Asp Gln Ile Phe Thr Ala Asp Asn Ser Thr Tyr Val Pro<br>1410 1415 1420      |      |      |
| 40 | Gln Gln Pro Ala Pro Ser Phe Glu Glu Met Ile Tyr Gln Phe Asn Asn<br>1425 1430 1435 1440 |      |      |
|    | Leu Thr Ile Asp Cys Lys Asn Leu Asn Phe Ile Asp Asn Gln Ala His<br>1445 1450 1455      |      |      |
| 45 | Ile Glu Ile Asp Phe Thr Ala Thr Ala Gln Asp Gly Arg Phe Leu Gly<br>1460 1465 1470      |      |      |
|    | Ala Glu Thr Phe Ile Ile Pro Val Thr Lys Lys Val Leu Gly Thr Glu<br>1475 1480 1485      |      |      |
| 50 | Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly Val Gln Tyr Met Gln<br>1490 1495 1500      |      |      |
| 55 | Ile Gly Ala Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Gln Gln Leu<br>1505 1510 1515 1520 |      |      |
|    | Val Ser Arg Ala Asn Arg Gly Ile Asp Ala Val Leu Ser Met Glu Thr<br>1525 1530 1535      |      |      |
| 60 | Gln Asn Ile Gln Glu Pro Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu<br>1540 1545 1550      |      |      |
|    | Val Leu Asp Lys Tyr Asp Glu Ser Ile His Gly Thr Asn Lys Ser Phe<br>1555 1560 1565      |      |      |
| 65 | Ala Ile Glu Tyr Val Asp Ile Phe Lys Glu Asn Asp Ser Phe Val Ile<br>1570 1575 1580      |      |      |
| 70 | Tyr Gln Gly Glu Leu Ser Glu Thr Ser Gln Thr Val Val Lys Val Phe<br>1585 1590 1595 1600 |      |      |

Leu Ser Tyr Phe Ile Glu Ala Thr Gly Asn Lys Asn His Leu Trp Val  
 1605 1610 1615  
 5 Arg Ala Lys Tyr Gln Lys Glu Thr Thr Asp Lys Ile Leu Phe Asp Arg  
 1620 1625 1630  
 Thr Asp Glu Lys Asp Pro His Gly Trp Phe Leu Ser Asp Asp His Lys  
 1635 1640 1645  
 10 Thr Phe Ser Gly Leu Ser Ser Ala Gln Ala Leu Lys Asn Asp Ser Glu  
 1650 1655 1660  
 Pro Met Asp Phe Ser Gly Ala Asn Ala Leu Tyr Phe Trp Glu Leu Phe  
 1665 1670 1675 1680  
 15 Tyr Tyr Thr Pro Met Met Met Ala His Arg Leu Leu Gln Glu Gln Asn  
 1685 1690 1695  
 Phe Asp Ala Ala Asn His Trp Phe Arg Tyr Val Trp Ser Pro Ser Gly  
 1700 1705 1710  
 Tyr Ile Val Asp Gly Lys Ile Ala Ile Tyr His Trp Asn Val Arg Pro  
 1715 1720 1725  
 25 Leu Glu Glu Asp Thr Ser Trp Asn Ala Gln Gln Leu Asp Ser Thr Asp  
 1730 1735 1740  
 Pro Asp Ala Val Ala Gln Asp Asp Pro Met His Tyr Lys Val Ala Thr  
 1745 1750 1755 1760  
 30 Phe Met Ala Thr Leu Asp Leu Leu Met Ala Arg Gly Asp Ala Ala Tyr  
 1765 1770 1775  
 Arg Gln Leu Glu Arg Asp Thr Leu Ala Glu Ala Lys Met Trp Tyr Thr  
 1780 1785 1790  
 Gln Ala Leu Asn Leu Leu Gly Asp Glu Pro Gln Val Met Leu Ser Thr  
 1795 1800 1805  
 40 Thr Trp Ala Asn Pro Thr Leu Gly Asn Ala Ala Ser Lys Thr Thr Gln  
 1810 1815 1820  
 Gln Val Arg Gln Gln Val Leu Thr Gln Leu Arg Leu Asn Ser Arg Val  
 1825 1830 1835 1840  
 45 Lys Thr Pro Leu  
 1844

50 (2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1722 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 55 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54 (TcbA<sub>III</sub> coding region):

CTA GGA ACA GCC AAT TCC CTG ACC GCT TTA TTC CTG CCG CAG GAA AAT 48  
 Leu Gly Thr Ala Asn Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn  
 1 5 10 15  
 65 AGC AAG CTC AAA GGC TAC TGG CGG ACA CTG GCG CAG CGT ATG TTT AAT 96  
 Ser Lys Leu Lys Gly Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn  
 20 25 30

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | TTA | CGT | CAT | AAT | CTG | TCG | ATT | GAC | GGC | CAG | CCG | CTC | TCC | TTG | CCG | CTG | 144 |
|    | Leu | Arg | His | Asn | Leu | Ser | Ile | Asp | Gly | Gln | Pro | Leu | Ser | Leu | Pro | Leu |     |
|    |     | 35  |     |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |
| 5  | TAT | GCT | AAA | CCG | GCT | GAT | CCA | AAA | GCT | TTA | CTG | AGT | GCG | GCG | GTT | TCA | 192 |
|    | Tyr | Ala | Lys | Pro | Ala | Asp | Pro | Lys | Ala | Leu | Leu | Ser | Ala | Ala | Val | Ser |     |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |     |
| 10 | GCT | TCT | CAA | GGG | GGA | GCC | GAC | TTG | CCG | AAG | GCG | CCG | CTG | ACT | ATT | CAC | 240 |
|    | Ala | Ser | Gln | Gly | Gly | Ala | Asp | Leu | Pro | Lys | Ala | Pro | Leu | Thr | Ile | His |     |
|    |     | 65  |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |     |
| 15 | CGC | TTC | CCT | CAA | ATG | CTA | GAA | GGG | GCA | CGG | GGC | TTG | GTT | AAC | CAG | CTT | 288 |
|    | Arg | Phe | Pro | Gln | Met | Leu | Glu | Gly | Ala | Arg | Gly | Leu | Val | Asn | Gln | Leu |     |
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |     |
| 20 | ATA | CAG | TTC | GGT | AGT | TCA | CTA | TTG | GGG | TAC | AGT | GAG | CGT | CAG | GAT | GCG | 336 |
|    | Ile | Gln | Phe | Gly | Ser | Ser | Leu | Leu | Gly | Tyr | Ser | Glu | Arg | Gln | Asp | Ala |     |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |     |
| 25 | GAA | GCT | ATG | AGT | CAA | CTA | CTG | CAA | ACC | CAA | GCC | AGC | GAG | TTA | ATA | CTG | 384 |
|    | Glu | Ala | Met | Ser | Gln | Leu | Leu | Gln | Thr | Gln | Ala | Ser | Glu | Leu | Ile | Leu |     |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
| 30 | ACC | AGT | ATT | CGT | ATG | CAG | GAT | AAC | CAA | TTG | GCA | GAG | CTG | GAT | TCG | GAA | 432 |
|    | Thr | Ser | Ile | Arg | Met | Gln | Asp | Asn | Gln | Leu | Ala | Glu | Leu | Asp | Ser | Glu |     |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |     |
| 35 | AAA | ACC | GCC | TTG | CAA | GTC | TCT | TTA | GCT | GGA | GTG | CAA | CAA | CGG | TTT | GAC | 480 |
|    | Lys | Thr | Ala | Leu | Gln | Val | Ser | Leu | Ala | Gly | Val | Gln | Gln | Arg | Phe | Asp |     |
|    |     | 145 |     |     |     | 150 |     |     |     | 155 |     |     |     |     |     | 160 |     |
| 40 | AGC | TAT | AGC | CAA | CTG | TAT | GAG | GAG | AAC | ATC | AAC | GCA | GGT | GAG | CAG | CGA | 528 |
|    | Ser | Tyr | Ser | Gln | Leu | Tyr | Glu | Glu | Asn | Ile | Asn | Ala | Gly | Glu | Gln | Arg |     |
|    |     |     |     |     | 165 |     |     |     | 170 |     |     |     |     |     | 175 |     |     |
| 45 | GCG | CTG | GCG | TTA | CGC | TCA | GAA | TCT | GCT | ATT | GAG | TCT | CAG | GGA | GCG | CAG | 576 |
|    | Ala | Leu | Ala | Leu | Arg | Ser | Glu | Ser | Ala | Ile | Glu | Ser | Gln | Gly | Ala | Gln |     |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |
| 50 | ATT | TCC | CGT | ATG | GCA | GGC | GCG | GGT | GTT | GAT | ATG | GCA | CCA | AAT | ATC | TTC | 624 |
|    | Ile | Ser | Arg | Met | Ala | Gly | Ala | Gly | Val | Asp | Met | Ala | Pro | Asn | Ile | Phe |     |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |
| 55 | GGC | CTG | GCT | GAT | GGC | GGC | ATG | CAT | TAT | GGT | GCT | ATT | GCC | TAT | GCC | ATC | 672 |
|    | Gly | Leu | Ala | Asp | Gly | Gly | Met | His | Tyr | Gly | Ala | Ile | Ala | Tyr | Ala | Ile |     |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |     |
| 60 | GCT | GAC | GGT | ATT | GAG | TTG | AGT | GCT | TCT | GCC | AAG | ATG | GTT | GAT | GCG | GAG | 720 |
|    | Ala | Asp | Gly | Ile | Glu | Leu | Ser | Ala | Ser | Ala | Lys | Met | Val | Asp | Ala | Glu |     |
|    |     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |     |
| 65 | AAA | GTT | GCT | CAG | TCG | GAA | ATA | TAT | CGC | CGT | CGC | CGT | CAA | GAA | TGG | AAA | 768 |
|    | Lys | Val | Ala | Gln | Ser | Glu | Ile | Tyr | Arg | Arg | Arg | Arg | Gln | Glu | Trp | Lys |     |
|    |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |     |
| 70 | ATT | CAG | CGT | GAC | AAC | GCA | CAA | GCG | GAG | ATT | AAC | CAG | TTA | AAC | GCG | CAA | 816 |
|    | Ile | Gln | Arg | Asp | Asn | Ala | Gln | Ala | Glu | Ile | Asn | Gln | Leu | Asn | Ala | Gln |     |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |     |
| 75 | CTG | GAA | TCA | CTG | TCT | ATT | CGC | CGT | GAA | GCC | GCT | GAA | ATG | CAA | AAA | GAG | 864 |
|    | Leu | Glu | Ser | Leu | Ser | Ile | Arg | Arg | Glu | Ala | Ala | Glu | Met | Gln | Lys | Glu |     |
|    |     |     |     | 275 |     |     |     | 280 |     |     |     |     | 285 |     |     |     |     |
| 80 | TAC | CTG | AAA | ACC | CAG | CAA | GCT | CAG | GCG | CAG | GCA | CAA | CTT | ACT | TTC | TTA | 912 |
|    | Tyr | Leu | Lys | Thr | Gln | Gln | Ala | Gln | Ala | Gln | Ala | Gln | Leu | Thr | Phe | Leu |     |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |     |
| 85 | AGA | AGC | AAA | TTC | AGT | AAT | CAA | GCG | TTA | TAT | AGT | TGG | TTA | CGA | GGG | CGT | 960 |
|    | Arg | Ser | Lys | Phe | Ser | Asn | Gln | Ala | Leu | Tyr | Ser | Trp | Leu | Arg | Gly | Arg |     |
|    |     | 305 |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |     |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | TTG | TCA | GGT | ATT | TAT | TTC | CAG | TTC | TAT | GAC | TTG | GCC | GTA | TCA | CGT | TGC | 1008 |
|    | Leu | Ser | Gly | Ile | Tyr | Phe | Gln | Phe | Tyr | Asp | Leu | Ala | Val | Ser | Arg | Cys |      |
|    |     |     |     | 325 |     |     |     |     |     | 330 |     |     |     |     | 335 |     |      |
| 5  | CTG | ATG | GCA | GAG | CAA | TCC | TAT | CAA | TGG | GAA | GCT | AAT | GAT | AAT | TCC | ATT | 1056 |
|    | Leu | Met | Ala | Glu | Gln | Ser | Tyr | Gln | Trp | Glu | Ala | Asn | Asp | Asn | S r | Ile |      |
|    |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |      |
| 10 | AGC | TTT | GTC | AAA | CCG | GGT | GCA | TGG | CAA | GGA | ACT | TAC | GCC | GGC | TTA | TTG | 1104 |
|    | Ser | Phe | Val | Lys | Pro | Gly | Ala | Trp | Gln | Gly | Thr | Tyr | Ala | Gly | Leu | Leu |      |
|    |     |     |     | 355 |     |     |     | 360 |     |     |     |     | 365 |     |     |     |      |
| 15 | TGT | GGA | GAA | GCT | TTG | ATA | CAA | AAT | CTG | GCA | CAA | ATG | GAA | GAG | GCA | TAT | 1152 |
|    | Cys | Gly | Glu | Ala | Leu | Ile | Gln | Asn | Leu | Ala | Gln | Met | Glu | Glu | Ala | Tyr |      |
|    |     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |      |
| 20 | CTG | AAA | TGG | GAA | TCT | CGC | GCT | TTG | GAA | GTA | GAA | CGC | ACG | GTT | TCA | TTG | 1200 |
|    | Leu | Lys | Trp | Glu | Ser | Arg | Ala | Leu | Glu | Val |     | Arg | Thr | Val | Ser | Leu |      |
|    |     | 385 |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |      |
| 25 | GCA | GTG | GTT | TAT | GAT | TCA | CTG | GAA | GGT | AAT | GAT | CGT | TTT | AAT | TTA | GCG | 1248 |
|    | Ala | Val | Val | Tyr | Asp | Ser | Leu | Glu | Gly | Asn | Asp | Arg | Phe | Asn | Leu | Ala |      |
|    |     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |      |
| 30 | GAA | CAA | ATA | CCT | GCA | TTA | TTG | GAT | AAG | GGG | GAG | GGA | ACA | GCA | GGA | ACT | 1296 |
|    | Glu | Gln | Ile | Pro | Ala | Leu | Leu | Asp | Lys | Gly | Glu | Gly | Thr | Ala | Gly | Thr |      |
|    |     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |      |
| 35 | AAA | GAA | AAT | GGG | TTA | TCA | TTG | GCT | AAT | GCT | ATC | CTG | TCA | GCT | TCG | GTC | 1344 |
|    | Lys | Glu | Asn | Gly | Leu | Ser | Leu | Ala | Asn | Ala | Ile | Leu | Ser | Ala | Ser | Val |      |
|    |     |     | 435 |     |     |     | 440 |     |     |     |     | 445 |     |     |     |     |      |
| 40 | AAA | TTG | TCC | GAC | TTG | AAA | CTG | GGA | ACG | GAT | TAT | CCA | GAC | AGT | ATC | GTT | 1392 |
|    | Lys | Leu | Ser | Asp | Leu | Lys | Leu | Gly | Thr | Asp | Tyr | Pro | Asp | Ser | Ile | Val |      |
|    |     | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |      |
| 45 | GGT | AGC | AAC | AAG | GTT | CGT | CGT | ATT | AAG | CAA | ATC | AGT | GTT | TCG | CTA | CCT | 1440 |
|    | Gly | Ser | Asn | Lys | Val | Arg | Arg | Ile | Lys | Gln | Ile | Ser | Val | Ser | Leu | Pro |      |
|    |     | 465 |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |      |
| 50 | GCA | TTG | GTT | GGG | CCT | TAT | CAG | GAT | GTT | CAG | GCT | ATG | CTC | AGC | TAT | GGT | 1488 |
|    | Ala | Leu | Val | Gly | Pro | Tyr | Gln | Asp | Val | Gln | Ala | Met | Leu | Ser | Tyr | Gly |      |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |      |
| 55 | GGC | AGT | ACT | CAA | TTG | CCG | AAA | GGT | TGT | TCA | GCG | TTG | GCT | GTG | TCT | CAT | 1536 |
|    | Gly | Ser | Thr | Gln | Leu | Pro | Lys | Gly | Cys | Ser | Ala | Leu | Ala | Val | Ser | His |      |
|    |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |      |
| 60 | GGT | ACC | AAT | GAT | AGT | GGT | CAG | TTC | CAG | TTG | GAT | TTC | AAT | GAC | GGC | AAA | 1584 |
|    | Gly | Thr | Asn | Asp | Ser | Gly | Gln | Phe | Gln | Leu | Asp | Phe | Asn | Asp | Gly | Lys |      |
|    |     |     | 515 |     |     |     | 520 |     |     |     |     | 525 |     |     |     |     |      |
| 65 | TAC | CTG | CCA | TTT | GAA | GGT | ATT | GCT | CTT | GAT | GAT | CAG | GGT | ACA | CTG | AAT | 1632 |
|    | Tyr | Leu | Pro | Phe | Glu | Gly | Ile | Ala | Leu | Asp | Asp | Gln | Gly | Thr | Leu | Asn |      |
|    |     | 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |      |
| 70 | CTT | CAA | TTT | CCG | AAT | GCT | ACC | GAC | AAG | CAG | AAA | GCA | ATA | TTG | CAA | ACT | 1680 |
|    | Leu | Gln | Phe | Pro | Asn | Ala | Thr | Asp | Lys | Gln | Lys | Ala | Ile | Leu | Gln | Thr |      |
|    |     | 545 |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |      |
|    | ATG | AGC | GAT | ATT | ATT | TTG | CAT | ATT | CGT | TAT | ACC | ATC | CGT | TAA |     |     | 1722 |
|    | Met | Ser | Asp | Ile | Ile | Leu | His | Ile | Arg | Tyr | Thr | Ile | Arg | ... |     |     |      |
|    |     |     |     |     | 565 |     |     |     | 570 |     |     | 573 |     |     |     |     |      |

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 amino acids

(B) TYPE: amino acids

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55 (TcbAiii):

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 10 | Leu | Gly | Thr | Ala | Asn | Ser | Leu | Thr | Ala | Leu | Phe | Leu | Pro | Gln | Glu | Asn |
|    | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
|    | Ser | Lys | Leu | Lys | Gly | Tyr | Trp | Arg | Thr | Leu | Ala | Gln | Arg | Met | Phe | Asn |
|    |     |     | 20  |     |     |     |     | 25  |     |     |     |     |     | 30  |     |     |
| 15 | Leu | Arg | His | Asn | Leu | Ser | Ile | Asp | Gly | Gln | Pro | Leu | Ser | Leu | Pro | Leu |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
|    | Tyr | Ala | Lys | Pro | Ala | Asp | Pro | Lys | Ala | Leu | Leu | Ser | Ala | Ala | Val | Ser |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| 20 | Ala | Ser | Gln | Gly | Gly | Ala | Asp | Leu | Pro | Lys | Ala | Pro | Leu | Thr | Ile | His |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
|    | Arg | Phe | Pro | Gln | Met | Leu | Glu | Gly | Ala | Arg | Gly | Leu | Val | Asn | Gln | Leu |
| 25 |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Ile | Gln | Phe | Gly | Ser | Ser | Leu | Leu | Gly | Tyr | Ser | Glu | Arg | Gln | Asp | Ala |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| 30 | Glu | Ala | Met | Ser | Gln | Leu | Leu | Gln | Thr | Gln | Ala | Ser | Glu | Leu | Ile | Leu |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|    | Thr | Ser | Ile | Arg | Met | Gln | Asp | Asn | Gln | Leu | Ala | Glu | Leu | Asp | Ser | Glu |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 35 | Lys | Thr | Ala | Leu | Gln | Val | Ser | Leu | Ala | Gly | Val | Gln | Gln | Arg | Phe | Asp |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Ser | Tyr | Ser | Gln | Leu | Tyr | Glu | Glu | Asn | Ile | Asn | Ala | Gly | Glu | Gln | Arg |
| 40 |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Ala | Leu | Ala | Leu | Arg | Ser | Glu | Ser | Ala | Ile | Glu | Ser | Gln | Gly | Ala | Gln |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| 45 | Ile | Ser | Arg | Met | Ala | Gly | Ala | Gly | Val | Asp | Met | Ala | Pro | Asn | Ile | Phe |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Gly | Leu | Ala | Asp | Gly | Gly | Met | His | Tyr | Gly | Ala | Ile | Ala | Tyr | Ala | Ile |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 50 | Ala | Asp | Gly | Ile | Glu | Leu | Ser | Ala | Ser | Ala | Lys | Met | Val | Asp | Ala | Glu |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
|    | Lys | Val | Ala | Gln | Ser | Glu | Ile | Tyr | Arg | Arg | Arg | Arg | Gln | Glu | Trp | Lys |
| 55 |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | Ile | Gln | Arg | Asp | Asn | Ala | Gln | Ala | Glu | Ile | Asn | Gln | Leu | Asn | Ala | Gln |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| 60 | Leu | Glu | Ser | Leu | Ser | Ile | Arg | Arg | Glu | Ala | Ala | Glu | Met | Gln | Lys | Glu |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
|    | Tyr | Leu | Lys | Thr | Gln | Gln | Ala | Gln | Ala | Gln | Ala | Gln | Leu | Thr | Phe | Leu |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| 65 | Arg | Ser | Lys | Phe | Ser | Asn | Gln | Ala | Leu | Tyr | Ser | Trp | Leu | Arg | Gly | Arg |
|    | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
|    | Leu | Ser | Gly | Ile | Tyr | Phe | Gln | Phe | Tyr | Asp | Leu | Ala | Val | Ser | Arg | Cys |
| 70 |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Leu | Met | Ala | Glu | Gln | Ser | Tyr | Gln | Trp | Glu | Ala | Asn | Asp | Asn | Ser | Ile |
|    |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| 5  | Ser | Phe | Val | Lys | Pro | Gly | Ala | Trp | Gln | Gly | Thr | Tyr | Ala | Gly | Leu | Leu |
|    |     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
|    | Cys | Gly | Glu | Ala | Leu | Ile | Gln | Asn | Leu | Ala | Gln | Met | Glu | Glu | Ala | Tyr |
| 10 |     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |
|    | Leu | Lys | Trp | Glu | Ser | Arg | Ala | Leu | Glu | Val | Glu | Arg | Thr | Val | Ser | Leu |
|    | 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |
| 15 | Ala | Val | Val | Tyr | Asp | Ser | Leu | Glu | Gly | Asn | Asp | Arg | Phe | Asn | Leu | Ala |
|    |     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |
|    | Glu | Gln | Ile | Pro | Ala | Leu | Leu | Asp | Lys | Gly | Glu | Gly | Thr | Ala | Gly | Thr |
|    |     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |
| 20 | Lys | Glu | Asn | Gly | Leu | Ser | Leu | Ala | Asn | Ala | Ile | Leu | Ser | Ala | Ser | Val |
|    |     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
|    | Lys | Leu | Ser | Asp | Leu | Lys | Leu | Gly | Thr | Asp | Tyr | Pro | Asp | Ser | Ile | Val |
| 25 |     | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |
|    | Gly | Ser | Asn | Lys | Val | Arg | Arg | Ile | Lys | Gln | Ile | Ser | Val | Ser | Leu | Pro |
|    | 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |
| 30 | Ala | Leu | Val | Gly | Pro | Tyr | Gln | Asp | Val | Gln | Ala | Met | Leu | Ser | Tyr | Gly |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |
|    | Gly | Ser | Thr | Gln | Leu | Pro | Lys | Gly | Cys | Ser | Ala | Leu | Ala | Val | Ser | His |
|    |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |
| 35 | Gly | Thr | Asn | Asp | Ser | Gly | Gln | Phe | Gln | Leu | Asp | Phe | Asn | Asp | Gly | Lys |
|    |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |
|    | Tyr | Leu | Pro | Phe | Glu | Gly | Ile | Ala | Leu | Asp | Asp | Gln | Gly | Thr | Leu | Asn |
| 40 |     | 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |
|    | Leu | Gln | Phe | Pro | Asn | Ala | Thr | Asp | Lys | Gln | Lys | Ala | Ile | Leu | Gln | Thr |
|    | 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |
| 45 | Met | Ser | Asp | Ile | Ile | Leu | His | Ile | Arg | Tyr | Thr | Ile | Arg | ... |     |     |
|    |     |     |     | 565 |     |     |     |     |     | 570 |     |     | 573 |     |     |     |

## (2) INFORMATION FOR SEQ ID NO:56

(i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 2898 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56 (tcca)

|    |    |                                                                 |     |
|----|----|-----------------------------------------------------------------|-----|
| 60 | 1  | ATG AAT CAA CTC GCC AGT CCC CTG ATT TCC CGC ACC GAA GAG ATC CAC | 48  |
|    | 1  | Met Asn Gln Leu Ala Ser Pro Leu Ile Ser Arg Thr Glu Glu Ile His | 16  |
|    | 49 | AAC TTA CCC GGT AAA TTG ACC GAT CTT GGT TAT ACC TCA GTG TTT GAT | 96  |
| 65 | 17 | Asn Leu Pro Gly Lys Leu Thr Asp Leu Gly Tyr Thr Ser Val Phe Asp | 32  |
|    | 97 | GTG GTA CGT ATG CCG CGT GAG CGT TTT ATT CGT GAG CAT CGT GCT GAT | 144 |



|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | 33  | Val | Val | Arg | Met | Pro | Arg | Glu | Arg | Phe | Ile | Arg | Glu | His | Arg | Ala | Asp | 43  |
|    | 145 | CTC | GGG | CGC | ACT | GCT | GAA | AAA | ATG | TAT | GAC | CTG | GCA | GTG | GGC | TAT | GCT | 192 |
| 5  | 49  | Leu | Gly | Arg | Ser | Ala | Glu | Lys | Met | Tyr | Asp | Leu | Ala | Val | Gly | Tyr | Ala | 64  |
|    | 193 | CAT | CAG | GTG | TTA | CAC | CAT | TTT | CGC | CGT | AAT | TCT | CTT | AGT | GAA | GCT | GTT | 240 |
| 10 | 65  | His | Gln | Val | Leu | His | His | Phe | Arg | Arg | Asn | Ser | Leu | Ser | Glu | Ala | Val | 30  |
|    | 241 | CAG | TTT | GGC | TTG | AGA | AGT | CCG | TTC | TCC | GTA | TCA | GGC | CCG | GAT | TAC | GCC | 288 |
|    | 81  | Gln | Phe | Gly | Leu | Arg | Ser | Pro | Phe | Ser | Val | Ser | Gly | Pro | Asp | Tyr | Ala | 96  |
| 15 | 289 | AAT | CAG | TTT | CTT | GAT | GCA | AAC | ACG | GGT | TGG | AAA | GAT | AAA | GCA | CCA | AGT | 336 |
|    | 97  | Asn | Gln | Phe | Leu | Asp | Ala | Asn | Thr | Gly | Trp | Lys | Asp | Lys | Ala | Pro | Ser | 112 |
| 20 | 337 | GGA | TCA | CCG | GAA | GCC | AAT | GAT | GCG | CCG | GTA | GCC | TAT | CTG | ACT | CAT | ATT | 384 |
|    | 113 | Gly | Ser | Pro | Glu | Ala | Asn | Asp | Ala | Pro | Val | Ala | Tyr | Leu | Thr | His | Ile | 128 |
| 25 | 385 | TAT | CAA | TTG | GCC | CTT | GAA | CAG | GAA | AAG | AAT | GGC | GCC | ACT | ACC | ATT | ATG | 432 |
|    | 129 | Tyr | Gln | Leu | Ala | Leu | Glu | Gln | Glu | Lys | Asn | Gly | Ala | Thr | Thr | Ile | Met | 144 |
| 30 | 433 | AAT | ACG | CTG | GCG | GAG | CGT | CGC | CCC | GAT | CTG | GGT | GCT | TTG | TTA | ATT | AAT | 480 |
|    | 145 | Asn | Thr | Leu | Ala | Glu | Arg | Arg | Pro | Asp | Leu | Gly | Ala | Leu | Leu | Ile | Asn | 160 |
|    | 481 | GAT | AAA | GCA | ATC | AAT | GAG | GTG | ATA | CCG | CAA | TTG | CAG | TTG | GTC | AAT | GAA | 528 |
|    | 161 | Asp | Lys | Ala | Ile | Asn | Glu | Val | Ile | Pro | Gln | Leu | Gln | Leu | Val | Asn | Glu | 176 |
| 35 | 529 | ATT | CTG | TCC | AAA | GCT | ATT | CAG | AAG | AAA | CTG | AGT | TTG | ACT | GAT | CTG | GAA | 576 |
|    | 177 | Ile | Leu | Ser | Lys | Ala | Ile | Gln | Lys | Lys | Leu | Ser | Leu | Thr | Asp | Leu | Glu | 192 |
| 40 | 577 | GCG | GTA | AAC | GCC | AGA | CTT | TCC | ACT | ACC | CGT | TAC | CCG | AAT | AAT | CTG | CCG | 624 |
|    | 193 | Ala | Val | Asn | Ala | Arg | Leu | Ser | Thr | Thr | Arg | Tyr | Pro | Asn | Asn | Leu | Pro | 208 |
| 45 | 625 | TAT | CAT | TAT | GGT | CAT | CAG | CAG | ATT | CAG | ACA | GCT | CAA | TCG | GTA | TTG | GGT | 672 |
|    | 209 | Tyr | His | Tyr | Gly | His | Gln | Gln | Ile | Gln | Thr | Ala | Gln | Ser | Val | Leu | Gly | 224 |
| 50 | 673 | ACT | ACG | TTG | CAA | GAT | ATC | ACT | TTG | CCA | CAG | ACG | CTG | GAT | CTG | CCG | CAA | 720 |
|    | 225 | Thr | Thr | Leu | Gln | Asp | Ile | Thr | Leu | Pro | Gln | Thr | Leu | Asp | Leu | Pro | Gln | 240 |
|    | 721 | AAC | TTC | TGG | GCA | ACA | GCA | AAA | GGA | AAA | CTG | AGC | GAT | ACG | ACT | GCC | AGT | 768 |
|    | 241 | Asn | Phe | Trp | Ala | Thr | Ala | Lys | Gly | Lys | Leu | Ser | Asp | Thr | Thr | Ala | Ser | 256 |
| 55 | 769 | GCT | TTG | ACC | CGA | CTG | CAA | ATC | ATG | GCG | AGT | CAG | TTT | TCG | CCA | GAG | CAG | 816 |
|    | 257 | Ala | Leu | Thr | Arg | Leu | Gln | Ile | Met | Ala | Ser | Gln | Phe | Ser | Pro | Glu | Gln | 272 |
| 60 | 817 | CAG | AAA | ATC | ATT | ACG | GAG | ACT | GTC | GGT | CAG | GAT | TTC | TAT | CAG | CTT | AAC | 864 |
|    | 273 | Gln | Lys | Ile | Ile | Thr | Glu | Thr | Val | Gly | Gln | Asp | Phe | Tyr | Gln | Leu | Asn | 288 |
| 65 | 865 | TAT | GGT | GAC | AGT | TCG | CTT | ACT | GTG | AAT | AGT | TTC | AGC | GAC | ATG | ACC | ATA | 912 |
|    | 289 | Tyr | Gly | Asp | Ser | Ser | Leu | Thr | Val | Asn | Ser | Phe | Ser | Asp | Met | Thr | Ile | 304 |

|    |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | 913  | ATG | ACT | GAT | CGA | ACA | AGT | TTG | ACT | GTA | CCC | CAG | GTA | GAA | CTG | ATG | TTG | 360  |
|    | 305  | Met | Thr | Asp | Arg | Thr | Ser | Leu | Thr | Val | Pro | Gln | Val | Glu | Leu | Met | Leu | 320  |
| 5  |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 961  | TGT | TCA | ACT | GTC | GGA | GGT | TCT | ACG | GTT | GTT | AAG | TCT | GAT | AAT | GTC | AGT | 1003 |
|    | 321  | Cys | Ser | Thr | Val | Gly | Gly | Ser | Thr | Val | Val | Lys | Ser | Asp | Asn | Val | Ser | 336  |
| 10 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1009 | TCT | GGT | GAC | ACG | ACA | GCG | ACG | CCA | TTT | GCG | TAT | GGC | GCC | CGC | TTT | ATT | 1056 |
|    | 337  | Ser | Gly | Asp | Thr | Thr | Ala | Thr | Pro | Phe | Ala | Tyr | Gly | Ala | Arg | Phe | Ile | 352  |
| 15 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1057 | CAT | GCC | GGT | AAG | CCG | GAG | GCG | ATT | ACC | CTG | AGT | CGC | AGT | GGT | GCG | GAG | 1104 |
|    | 353  | His | Ala | Gly | Lys | Pro | Glu | Ala | Ile | Thr | Leu | Ser | Arg | Ser | Gly | Ala | Glu | 368  |
| 20 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1105 | GCG | CAT | TTT | GCT | CTG | ACG | GTT | AAC | AAT | CTG | ACA | GAT | GAC | AAG | TTG | GAC | 1152 |
|    | 369  | Ala | His | Phe | Ala | Leu | Thr | Val | Asn | Asn | Leu | Thr | Asp | Asp | Lys | Leu | Asp | 384  |
| 25 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1153 | CGT | ATT | AAC | CGC | ACA | GTG | CGC | CTG | CAA | AAA | TGG | CTG | AAT | CTG | CCT | TAT | 1200 |
|    | 385  | Arg | Ile | Asn | Arg | Thr | Val | Arg | Leu | Gln | Lys | Trp | Leu | Asn | Leu | Pro | Tyr | 400  |
| 30 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1201 | GAG | GAT | ATT | GAC | CTG | TTA | GTG | ACT | TCT | GCT | ATG | GAT | GCG | GAA | ACA | GGA | 1248 |
|    | 401  | Glu | Asp | Ile | Asp | Leu | Leu | Val | Thr | Ser | Ala | Met | Asp | Ala | Glu | Thr | Gly | 416  |
| 35 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1249 | AAT | ACC | GCG | CTG | TCG | ATG | AAC | GAC | AAT | ACG | CTG | CGT | ATG | TTG | GGA | GTC | 1296 |
|    | 417  | Asn | Thr | Ala | Leu | Ser | Met | Asn | Asp | Asn | Thr | Leu | Arg | Met | Leu | Gly | Val | 432  |
| 40 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1297 | TTC | AAA | CAT | TAT | CAG | GCG | AAG | TAT | GGT | GTT | AGC | GCT | AAA | CAA | TTT | GCT | 1344 |
|    | 433  | Phe | Lys | His | Tyr | Gln | Ala | Lys | Tyr | Gly | Val | Ser | Ala | Lys | Gln | Phe | Ala | 448  |
| 45 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1345 | GGC | TGG | CTG | CGC | GTA | GTG | GCC | CCG | TTT | GCC | ATT | ACA | CCG | GCA | ACG | CCG | 1392 |
|    | 449  | Gly | Trp | Leu | Arg | Val | Val | Ala | Pro | Phe | Ala | Ile | Thr | Pro | Ala | Thr | Pro | 464  |
| 50 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1393 | TTT | TTA | GAC | CAA | GTG | TTT | AAC | TCC | GTC | GGC | ACC | TTT | GAT | ACA | CCG | TTT | 1440 |
|    | 465  | Phe | Leu | Asp | Gln | Val | Phe | Asn | Ser | Val | Gly | Thr | Phe | Asp | Thr | Pro | Phe | 480  |
| 55 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1441 | GTG | ATA | GAT | AAT | CAG | GAT | TTT | GTC | TAT | ACA | TTG | ACC | ACC | GGG | GGC | GAT | 1488 |
|    | 481  | Val | Ile | Asp | Asn | Gln | Asp | Phe | Val | Tyr | Thr | Leu | Thr | Thr | Gly | Gly | Asp | 496  |
| 60 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1489 | GGG | GCG | CGT | GTT | AAG | CAT | ATC | AGC | ACG | GCA | CTG | GGC | CTC | AAT | CAT | CGT | 1536 |
|    | 497  | Gly | Ala | Arg | Val | Lys | His | Ile | Ser | Thr | Ala | Leu | Gly | Leu | Asn | His | Arg | 512  |
| 65 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1537 | CAG | TTC | CTG | TTA | TTG | GCG | GAT | AAT | ATT | GCC | CGT | CAA | CAG | GGG | AAT | GTC | 1584 |
|    | 513  | Gln | Phe | Leu | Leu | Leu | Ala | Asp | Asn | Ile | Ala | Arg | Gln | Gln | Gly | Asn | Val | 528  |
| 70 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1585 | ACG | CAA | AGC | ACA | CTC | AAC | TGT | AAT | CTG | TTT | GTG | GTG | TCA | GCT | TTC | TAC | 1632 |
|    | 529  | Thr | Gln | Ser | Thr | Leu | Asn | Cys | Asn | Leu | Phe | Val | Val | Ser | Ala | Phe | Tyr | 544  |
| 75 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1633 | CGT | CTG | GCT | AAT | TTG | GCG | CGC | ACA | TTG | GGG | ATA | AAT | CCA | GAG | TCT | TTC | 1680 |
|    | 545  | Arg | Leu | Ala | Asn | Leu | Ala | Arg | Thr | Leu | Gly | Ile | Asn | Pro | Glu | Ser | Phe | 560  |
| 80 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|    | 1681 | TGT | GCC | TTG | GTT | GAT | CGA | TTA | GAT | GCA | GGT | ACA | GGC | ATC | GTC | TGG | CAG | 1728 |

|    |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | 561  | Cys | Ala | Leu | Val | Asp | Arg | Leu | Asp | Ala | Gly | Thr | Gly | Ile | Val | Trp | Gln | 579  |
| 5  | 1729 | CAA | TTG | GCA | GGG | AAA | CCC | ACA | ATC | ACG | GTA | CCA | CAA | AAA | GAT | TCC | CCG | 1775 |
|    | 577  | Gln | Leu | Ala | Gly | Lys | Pro | Thr | Ile | Thr | Val | Pro | Gln | Lys | Asp | Ser | Pro | 590  |
|    | 1777 | CTG | GCG | GCG | GAT | ATT | CTG | AGT | TTG | CTG | CAA | GCG | CTA | AGT | GCG | ATT | GCT |      |
|    | 1824 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
| 10 | 593  | Leu | Ala | Ala | Asp | Ile | Leu | Ser | Leu | Leu | Gln | Ala | Leu | Ser | Ala | Ile | Ala | 608  |
|    | 1825 | CAA | TGG | CAA | CAA | CAG | CAC | GAT | TTA | GAA | TTT | TCA | GCA | CTG | CTT | TTG | CTG | 1872 |
| 15 | 609  | Gln | Trp | Gln | Gln | Gln | His | Asp | Leu | Glu | Phe | Ser | Ala | Leu | Leu | Leu | Leu | 624  |
|    | 1873 | TTG | AGT | GAC | AAC | CCT | ATT | TCT | ACC | TCG | CAG | GGC | ACT | GAC | GAT | CAA | TTG | 1920 |
| 20 | 625  | Leu | Ser | Asp | Asn | Pro | Ile | Ser | Thr | Ser | Gln | Gly | Thr | Asp | Asp | Gln | Leu | 640  |
|    | 1921 | AAC | TTT | ATC | CGT | CAA | GTG | TGG | CAG | AAC | CTA | GGC | AGT | ACG | TTT | GTG | GGT | 1958 |
|    | 641  | Asn | Phe | Ile | Arg | Gln | Val | Trp | Gln | Asn | Leu | Gly | Ser | Thr | Phe | Val | Gly | 656  |
| 25 | 1969 | GCA | ACA | TTG | TTG | TCC | CGC | AGT | GGG | GCA | CCA | TTA | GTC | GAT | ACC | AAC | GGC | 2016 |
|    | 657  | Ala | Thr | Leu | Leu | Ser | Arg | Ser | Gly | Ala | Pro | Leu | Val | Asp | Thr | Asn | Gly | 672  |
| 30 | 2017 | CAC | GCT | ATT | GAC | TGG | TTT | GCT | CTG | CTC | TCA | GCA | GGT | AAT | AGT | CCG | CTT | 2064 |
|    | 673  | His | Ala | Ile | Asp | Trp | Phe | Ala | Leu | Leu | Ser | Ala | Gly | Asn | Ser | Pro | Leu | 688  |
| 35 | 2065 | ATC | GAT | AAG | GTT | GGT | CTG | GTG | ACT | GAT | GCT | GGC | ATA | CAA | AGT | GTT | ATA | 2112 |
|    | 689  | Ile | Asp | Lys | Val | Gly | Leu | Val | Thr | Asp | Ala | Gly | Ile | Gln | Ser | Val | Ile | 704  |
|    | 2113 | GCA | ACG | GTG | GTC | AAT | ACA | CAA | AGC | TTA | TCT | GAT | GAA | GAT | AAG | AAG | CTG | 2160 |
| 40 | 705  | Ala | Thr | Val | Val | Asn | Thr | Gln | Ser | Leu | Ser | Asp | Glu | Asp | Lys | Lys | Leu | 720  |
|    | 2161 | GCA | ATC | ACT | ACT | CTG | ACT | AAT | ACG | TTG | AAT | CAG | GTA | CAG | AAA | ACT | CAA | 2208 |
|    | 721  | Ala | Ile | Thr | Thr | Leu | Thr | Asn | Thr | Leu | Asn | Gln | Val | Gln | Lys | Thr | Gln | 736  |
| 45 | 2209 | CAG | GGC | GTG | GCC | GTC | AGT | CTG | TTG | GCG | CAG | ACT | CTG | AAC | GTG | AGT | CAG | 2256 |
|    | 737  | Gln | Gly | Val | Ala | Val | Ser | Leu | Leu | Ala | Gln | Thr | Leu | Asn | Val | Ser | Gln | 752  |
| 50 | 2257 | TCA | CTG | CCT | GCG | TTA | TTG | TTG | CGC | TGG | AGT | GGA | CAA | ACA | ACC | TAC | CAG | 2304 |
|    | 753  | Ser | Leu | Pro | Ala | Leu | Leu | Leu | Arg | Trp | Ser | Gly | Gln | Thr | Thr | Tyr | Gln | 763  |
| 55 | 2305 | TGG | TTG | AGT | GCG | ACT | TGG | GCA | TTG | AAG | GAT | GCC | GTT | AAG | ACT | GCC | GCC | 2352 |
|    | 769  | Trp | Leu | Ser | Ala | Thr | Trp | Ala | Leu | Lys | Asp | Ala | Val | Lys | Thr | Ala | Ala | 784  |
|    | 2353 | GAT | ATT | CCC | GCT | GAC | TAT | CTG | CGT | CAA | TTA | CGT | GAA | GTG | GTA | CGC | CGC | 2400 |
| 60 | 785  | Asp | Ile | Pro | Ala | Asp | Tyr | Leu | Arg | Gln | Leu | Arg | Glu | Val | Val | Arg | Arg | 800  |
|    | 2401 | TCC | TTG | TTG | ACC | CAA | CAA | TTC | ACG | CTG | AGT | CCT | GCA | ATG | GTG | CAA | ACC | 2448 |
|    | 801  | Ser | Leu | Leu | Thr | Gln | Gln | Phe | Thr | Leu | Ser | Pro | Ala | Met | Val | Gln | Thr | 816  |
| 65 |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |

|    |          |                                                                 |      |
|----|----------|-----------------------------------------------------------------|------|
|    | 2449     | TTG CTG GAC TAT CCA GCC TAT TTT GGC GCT TCC GCA GAA ACA GTG ACC | 2496 |
|    | 317      | Leu Leu Asp Tyr Pro Ala Tyr Phe Gly Ala Ser Ala Glu Thr Val Thr | 332  |
| 5  | 2497     | GAT ATC AGT TTG TGG ATG CTT TAT ACC CTG AGC TGT TAT AGC GAT TTA | 2544 |
|    | 333      | Asp Ile Ser Leu Trp Met Leu Tyr Thr Leu Ser Cys Tyr Ser Asp Leu | 348  |
| 10 | 2545     | TTG CTC CAA ATG GGT GAA GCT GGT GGT ACC GAA GAT GAT GTA CTG GCC | 2592 |
|    | 349      | Leu Leu Gln Met Gly Glu Ala Gly Gly Thr Glu Asp Asp Val Leu Ala | 364  |
| 15 | 2593     | TAC TTA CGC ACA GCT AAT GCT ACC ACA CCG TTG AGC CAA TCT GAT GCT | 2640 |
|    | 365      | Tyr Leu Arg Thr Ala Asn Ala Thr Thr Pro Leu Ser Gln Ser Asp Ala | 380  |
|    | 2641     | GCA CAG ACG TTG GCA ACG CTA TTG GGT TGG GAG GTT AAC GAG TTG CAA | 2688 |
|    | 381      | Ala Gln Thr Leu Ala Thr Leu Leu Gly Trp Glu Val Asn Glu Leu Gln | 396  |
| 20 | 2689     | GCC GCT TGG TCG GTA TTG GGC GGG ATT GCC AAA ACC ACA CCG CAA CTG | 2736 |
|    | 397      | Ala Ala Trp Ser Val Leu Gly Gly Ile Ala Lys Thr Thr Pro Gln Leu | 312  |
| 25 | 2737     | GAT GCG CTT CTG CGT TTG CAA CAG GCA CAG AAC CAA ACT GGT CTT GGC | 2784 |
|    | 413      | Asp Ala Leu Leu Arg Leu Gln Gln Ala Gln Asn Gln Thr Gly Leu Gly | 328  |
| 30 | 2785     | GTT ACA CAG CAA CAG CAA GGC TAT CTC CTG AGT CGT GAC AGT GAT TAT | 2832 |
|    | 429      | Val Thr Gln Gln Gln Gln Gly Tyr Leu Leu Ser Arg Asp Ser Asp Tyr | 344  |
| 35 | 2833     | ACC CTT TGG CAA AGC ACC GGT CAG GCG CTG GTG GCT GGC GTA TCC CAT | 2880 |
|    | 445      | Thr Leu Trp Gln Ser Thr Gly Gln Ala Leu Val Ala Gly Val Ser His | 360  |
|    | 2881     | GTC AAG GGC AGT AAC TGA                                         | 2898 |
|    | 461      | Val Lys Gly Ser Asn End                                         | 366  |
| 40 | (2)      | INFORMATION FOR SEQ ID NO:57                                    |      |
|    | (i)      | SEQUENCE CHARACTERISTICS:                                       |      |
|    | (A)      | LENGTH: 965 amino acids                                         |      |
|    | (B)      | TYPE: amino acid                                                |      |
| 45 | (C)      | TOPOLOGY: linear                                                |      |
|    | (ii)     | MOLECULE TYPE: protein                                          |      |
| 50 | (xi)     | SEQUENCE DESCRIPTION: SEQ ID NO:57 (TCCA peptide)               |      |
|    | Features | From To Description                                             |      |
|    |          | 1 10 SEQ ID NO:8                                                |      |
| 55 | 1        | Met Asn Gln Leu Ala Ser Pro Leu Ile Ser Arg Thr Glu Glu Ile His | 16   |
|    | 17       | Asn Leu Pro Gly Lys Leu Thr Asp Leu Gly Tyr Thr Ser Val Phe Asp | 32   |
|    | 33       | Val Val Arg Met Pro Arg Glu Arg Phe Ile Arg Glu His Arg Ala Asp | 48   |
| 60 | 49       | Leu Gly Arg Ser Ala Glu Lys Met Tyr Asp Leu Ala Val Gly Tyr Ala | 64   |
|    | 65       | His Gln Val Leu His His Phe Arg Arg Asn S r Leu Ser Glu Ala Val | 30   |
| 65 | 81       | Gln Phe Gly Leu Arg Ser Pro Phe Ser Val S r Gly Pro Asp Tyr Ala | 96   |

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | 97  | Asn | Gln | Phe | Leu | Asp | Ala | Asn | Thr | Gly | Trp | Lys | Asp | Lys | Ala | Pro | Ser | 111 |
|    | 113 | Gly | Ser | Pro | Glu | Ala | Asn | Asp | Ala | Pro | Val | Ala | Tyr | Leu | Thr | His | Ile | 128 |
| 5  | 129 | Tyr | Gln | Leu | Ala | Leu | Glu | Gln | Glu | Lys | Asn | Gly | Ala | Thr | Thr | Ile | Met | 144 |
|    | 145 | Asn | Thr | Leu | Ala | Glu | Arg | Arg | Pro | Asp | Leu | Gly | Ala | Leu | Leu | Ile | Asn | 160 |
|    | 161 | Asp | Lys | Ala | Ile | Asn | Glu | Val | Ile | Pro | Gln | Leu | Gln | Leu | Val | Asn | Glu | 176 |
| 10 | 177 | Ile | Leu | Ser | Lys | Ala | Ile | Gln | Lys | Lys | Leu | Ser | Leu | Thr | Asp | Leu | Glu | 192 |
|    | 193 | Ala | Val | Asn | Ala | Arg | Leu | Ser | Thr | Thr | Arg | Tyr | Pro | Asn | Asn | Leu | Pro | 208 |
| 15 | 209 | Tyr | His | Tyr | Gly | His | Gln | Gln | Ile | Gln | Thr | Ala | Gln | Ser | Val | Leu | Gly | 224 |
|    | 225 | Thr | Thr | Leu | Gln | Asp | Ile | Thr | Leu | Pro | Gln | Thr | Leu | Asp | Leu | Pro | Gln | 240 |
|    | 241 | Asn | Phe | Trp | Ala | Thr | Ala | Lys | Gly | Lys | Leu | Ser | Asp | Thr | Thr | Ala | Ser | 256 |
| 20 | 257 | Ala | Leu | Thr | Arg | Leu | Gln | Ile | Met | Ala | Ser | Gln | Phe | Ser | Pro | Glu | Gln | 272 |
|    | 273 | Gln | Lys | Ile | Ile | Thr | Glu | Thr | Val | Gly | Gln | Asp | Phe | Tyr | Gln | Leu | Asn | 288 |
| 25 | 289 | Tyr | Gly | Asp | Ser | Ser | Leu | Thr | Val | Asn | Ser | Phe | Ser | Asp | Met | Thr | Ile | 304 |
|    | 305 | Met | Thr | Asp | Arg | Thr | Ser | Leu | Thr | Val | Pro | Gln | Val | Glu | Leu | Met | Leu | 320 |
|    | 321 | Cys | Ser | Thr | Val | Gly | Gly | Ser | Thr | Val | Val | Lys | Ser | Asp | Asn | Val | Ser | 336 |
| 30 | 337 | Ser | Gly | Asp | Thr | Thr | Ala | Thr | Pro | Phe | Ala | Tyr | Gly | Ala | Arg | Phe | Ile | 352 |
|    | 353 | His | Ala | Gly | Lys | Pro | Glu | Ala | Ile | Thr | Leu | Ser | Arg | Ser | Gly | Ala | Glu | 368 |
| 35 | 369 | Ala | His | Phe | Ala | Leu | Thr | Val | Asn | Asn | Leu | Thr | Asp | Asp | Lys | Leu | Asp | 384 |
|    | 385 | Arg | Ile | Asn | Arg | Thr | Val | Arg | Leu | Gln | Lys | Trp | Leu | Asn | Leu | Pro | Tyr | 400 |
|    | 401 | Glu | Asp | Ile | Asp | Leu | Leu | Val | Thr | Ser | Ala | Met | Asp | Ala | Glu | Thr | Gly | 416 |
| 40 | 417 | Asn | Thr | Ala | Leu | Ser | Met | Asn | Asp | Asn | Thr | Leu | Arg | Met | Leu | Gly | Val | 432 |
|    | 433 | Phe | Lys | His | Tyr | Gln | Ala | Lys | Tyr | Gly | Val | Ser | Ala | Lys | Gln | Phe | Ala | 448 |
| 45 | 449 | Gly | Trp | Leu | Arg | Val | Val | Ala | Pro | Phe | Ala | Ile | Thr | Pro | Ala | Thr | Pro | 464 |
|    | 465 | Phe | Leu | Asp | Gln | Val | Phe | Asn | Ser | Val | Gly | Thr | Phe | Asp | Thr | Pro | Phe | 480 |
|    | 481 | Val | Ile | Asp | Asn | Gln | Asp | Phe | Val | Tyr | Thr | Leu | Thr | Thr | Gly | Gly | Asp | 496 |
| 50 | 497 | Gly | Ala | Arg | Val | Lys | His | Ile | Ser | Thr | Ala | Leu | Gly | Leu | Asn | His | Arg | 512 |
|    | 513 | Gln | Phe | Leu | Leu | Leu | Ala | Asp | Asn | Ile | Ala | Arg | Gln | Gln | Gly | Asn | Val | 528 |
| 55 | 529 | Thr | Gln | Ser | Thr | Leu | Asn | Cys | Asn | Leu | Phe | Val | Val | Ser | Ala | Phe | Tyr | 544 |
|    | 545 | Arg | Leu | Ala | Asn | Leu | Ala | Arg | Thr | Leu | Gly | Ile | Asn | Pro | Glu | Ser | Phe | 560 |
|    | 561 | Cys | Ala | Leu | Val | Asp | Arg | Leu | Asp | Ala | Gly | Thr | Gly | Ile | Val | Trp | Gln | 576 |
| 60 | 577 | Gln | Leu | Ala | Gly | Lys | Pro | Thr | Ile | Thr | Val | Pro | Gln | Lys | Asp | Ser | Pro | 592 |
|    | 593 | Leu | Ala | Ala | Asp | Ile | Leu | Ser | Leu | Leu | Gln | Ala | Leu | Ser | Ala | Ile | Ala | 608 |
| 65 | 609 | Gln | Trp | Gln | Gln | Gln | His | Asp | Leu | Glu | Phe | Ser | Ala | Leu | Leu | Leu | Leu | 624 |

625 Leu Ser Asp Asn Pro Ile Ser Thr Ser Gln Gly Thr Asp Asp Gln Leu 640  
 641 Asn Phe Ile Arg Gln Val Trp Gln Asn Leu Gly Ser Thr Phe Val Gly 656  
 5 657 Ala Thr Leu Leu Ser Arg Ser Gly Ala Pro Leu Val Asp Thr Asn Gly 672  
 673 His Ala Ile Asp Trp Phe Ala Leu Leu Ser Ala Gly Asn Ser Pro Leu 688  
 10 689 Ile Asp Lys Val Gly Leu Val Thr Asp Ala Gly Ile Gln Ser Val Ile 704  
 705 Ala Thr Val Val Asn Thr Gln Ser Leu Ser Asp Glu Asp Lys Lys Leu 720  
 721 Ala Ile Thr Thr Leu Thr Asn Thr Leu Asn Gln Val Gln Lys Thr Gln 736  
 15 737 Gln Gly Val Ala Val Ser Leu Leu Ala Gln Thr Leu Asn Val Ser Gln 752  
 753 Ser Leu Pro Ala Leu Leu Leu Arg Trp Ser Gly Gln Thr Thr Tyr Gln 768  
 20 769 Trp Leu Ser Ala Thr Trp Ala Leu Lys Asp Ala Val Lys Thr Ala Ala 784  
 785 Asp Ile Pro Ala Asp Tyr Leu Arg Gln Leu Arg Glu Val Val Arg Arg 800  
 801 Ser Leu Leu Thr Gln Gln Phe Thr Leu Ser Pro Ala Met Val Gln Thr 816  
 25 817 Leu Leu Asp Tyr Pro Ala Tyr Phe Gly Ala Ser Ala Glu Thr Val Thr 832  
 833 Asp Ile Ser Leu Trp Met Leu Tyr Thr Leu Ser Cys Tyr Ser Asp Leu 848  
 30 849 Leu Leu Gln Met Gly Glu Ala Gly Gly Thr Glu Asp Asp Val Leu Ala 864  
 865 Tyr Leu Arg Thr Ala Asn Ala Thr Thr Pro Leu Ser Gln Ser Asp Ala 880  
 881 Ala Gln Thr Leu Ala Thr Leu Leu Gly Trp Glu Val Asn Glu Leu Gln 896  
 35 897 Ala Ala Trp Ser Val Leu Gly Gly Ile Ala Lys Thr Thr Pro Gln Leu 912  
 913 Asp Ala Leu Leu Arg Leu Gln Gln Ala Gln Asn Gln Thr Gly Leu Gly 928  
 40 929 Val Thr Gln Gln Gln Gln Gly Tyr Leu Leu Ser Arg Asp Ser Asp Tyr 944  
 945 Thr Leu Trp Gln Ser Thr Gly Gln Ala Leu Val Ala Gly Val Ser His 960  
 961 Val Lys Gly Ser Asn 965

45

## (2) INFORMATION FOR SEQ ID NO:58

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4698 base pairs  
 (B) TYPE: nucleic acid  
 50 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58 (cccB)

1 ATG TTA TCG ACA ATG GAA AAA CAA CTG AAT GAA TCC CAG CGT GAT GCG 43  
 1 Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu Ser Gln Arg Asp Ala 16  
 60  
 49 TTG GTG ACT GGC TAT ATG AAT TTT GTG GCG CCG ACG TTG AAA GGC GTC 96  
 17 Leu Val Thr Gly Tyr Met Asn Phe Val Ala Pro Thr Leu Lys Gly Val 32

65

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | 97  | AGT | GGT | CAG | CCG | GTG | ACG | GTG | GAA | GAT | TTA | TAC | GAA | TAT | TTG | CTG | ATT | 144 |
|    | 33  | Ser | Gly | Gln | Pro | Val | Thr | Val | Glu | Asp | Leu | Tyr | Glu | Tyr | Leu | Leu | Ile | 48  |
| 5  | 145 | GAC | CCG | GAA | GTG | GCT | GAT | GAG | GTT | GAG | ACG | AGT | CGG | GTA | GCA | CAA | GCG | 192 |
|    | 49  | Asp | Pro | Glu | Val | Ala | Asp | Glu | Val | Glu | Thr | Ser | Arg | Val | Ala | Gln | Ala | 64  |
| 10 | 193 | ATT | GCC | AGC | ATA | CAG | CAA | TAT | ATG | ACT | CGT | CTG | GTC | AAC | GGC | TCT | GAA | 240 |
|    | 65  | Ile | Ala | Ser | Ile | Gln | Gln | Tyr | Met | Thr | Arg | Leu | Val | Asn | Gly | Ser | Glu | 80  |
| 15 | 241 | CCG | GGG | CGT | CAG | GCG | ATG | GAG | CCT | TCT | ACA | GCT | AAC | GAA | TGG | CGT | GAT | 288 |
|    | 81  | Pro | Gly | Arg | Gln | Ala | Met | Glu | Pro | Ser | Thr | Ala | Asn | Glu | Trp | Arg | Asp | 96  |
|    | 289 | AAT | GAT | AAC | CAA | TAT | GCT | ATC | TGG | GCT | GCG | GGG | GCT | GAG | GTT | CGA | AAT | 336 |
|    | 97  | Asn | Asp | Asn | Gln | Tyr | Ala | Ile | Trp | Ala | Ala | Gly | Ala | Glu | Val | Arg | Asn | 112 |
| 20 | 337 | TAC | GCT | GAA | AAC | TAT | ATT | TCA | CCC | ATC | ACC | CGG | CAG | GAA | AAA | AGC | CAT | 384 |
|    | 113 | Tyr | Ala | Glu | Asn | Tyr | Ile | Ser | Pro | Ile | Thr | Arg | Gln | Glu | Lys | Ser | His | 128 |
| 25 | 385 | TAT | TTC | TCG | GAG | CTG | GAG | ACG | ACT | TTA | AAT | CAG | AAT | CGA | CTC | GAT | CCG | 432 |
|    | 129 | Tyr | Phe | Ser | Glu | Leu | Glu | Thr | Thr | Leu | Asn | Gln | Asn | Arg | Leu | Asp | Pro | 144 |
| 30 | 433 | GAT | CGT | GTG | CAG | GAT | GCT | GTT | TTG | GCG | TAT | CTC | AAT | GAG | TTT | GAG | GCA | 480 |
|    | 145 | Asp | Arg | Val | Gln | Asp | Ala | Val | Leu | Ala | Tyr | Leu | Asn | Glu | Phe | Glu | Ala | 160 |
| 35 | 481 | GTG | AGT | AAT | CTA | TAT | GTG | CTC | AGT | GGT | TAT | ATT | AAT | CAG | GAT | AAA | TTT | 528 |
|    | 161 | Val | Ser | Asn | Leu | Tyr | Val | Leu | Ser | Gly | Tyr | Ile | Asn | Gln | Asp | Lys | Phe | 176 |
|    | 529 | GAC | CAA | GCT | ATC | TAC | TAC | TTT | ATT | GGT | CGC | ACT | ACC | ACT | AAA | CCG | TAT | 576 |
|    | 177 | Asp | Gln | Ala | Ile | Tyr | Tyr | Phe | Ile | Gly | Arg | Thr | Thr | Thr | Lys | Pro | Tyr | 192 |
| 40 | 577 | CGC | TAC | TAC | TGG | CGT | CAG | ATG | GAT | TTG | AGT | AAG | AAC | CGT | CAA | GAT | CCG | 624 |
|    | 193 | Arg | Tyr | Tyr | Trp | Arg | Gln | Met | Asp | Leu | Ser | Lys | Asn | Arg | Gln | Asp | Pro | 208 |
| 45 | 625 | GCA | GGG | AAT | CCG | GTG | ACG | CCA | AAT | TGC | TGG | AAT | GAT | TGG | CAG | GAA | ATC | 672 |
|    | 209 | Ala | Gly | Asn | Pro | Val | Thr | Pro | Asn | Cys | Trp | Asn | Asp | Trp | Gln | Glu | Ile | 224 |
| 50 | 673 | ACT | TTG | CCG | CTG | TCT | GGT | GAT | ACG | GTG | CTG | GAG | CAT | ACA | GTT | CGC | CCG | 720 |
|    | 225 | Thr | Leu | Pro | Leu | Ser | Gly | Asp | Thr | Val | Leu | Glu | His | Thr | Val | Arg | Pro | 240 |
| 55 | 721 | GTA | TTT | TAT | AAT | GAT | CGA | CTA | TAT | GTG | GCT | TGG | GTT | GAG | CGT | GAC | CCG | 768 |
|    | 241 | Val | Phe | Tyr | Asn | Asp | Arg | Leu | Tyr | Val | Ala | Trp | Val | Glu | Arg | Asp | Pro | 256 |
|    | 769 | GCA | GTA | CAG | AAG | GAT | GCT | GAC | GGT | AAA | AAC | ATC | GGT | AAA | ACC | CAT | GCC | 816 |
|    | 257 | Ala | Val | Gln | Lys | Asp | Ala | Asp | Gly | Lys | Asn | Ile | Gly | Lys | Thr | His | Ala | 272 |
| 60 | 817 | TAC | AAC | ATA | AAG | TTT | GGT | TAT | AAA | CGT | TAT | GAT | GAT | ACT | TGG | ACA | GCG | 864 |
|    | 273 | Tyr | Asn | Ile | Lys | Phe | Gly | Tyr | Lys | Arg | Tyr | Asp | Asp | Thr | Trp | Thr | Ala | 288 |
| 65 | 865 | CCG | AAT | ACG | ACC | ACG | TTA | ATG | ACA | CAA | CAA | GCA | GGG | GAA | AGT | TCA | GAA | 912 |
|    | 289 | Pro | Asn | Thr | Thr | Thr | Leu | Met | Thr | Gln | Gln | Ala | Gly | Glu | Ser | Ser | Glu | 304 |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
| 5  | 913  | ACA CAG CGA TCC AGC CTG CTG ATT GAT GAA TCT AGC ACC ACA TTG GGC | 960  |
|    | 305  | Thr Gln Arg Ser Ser Leu Leu Ile Asp Glu Ser Ser Thr Thr Leu Arg | 320  |
| 10 | 961  | CAA GTT AAT CTG TTG GCT ACC ACC GAT TTT AGT ATC GAT CCG ACG GAG | 1008 |
|    | 321  | Gln Val Asn Leu Leu Ala Thr Thr Asp Phe Ser Ile Asp Pro Thr Glu | 336  |
| 15 | 1009 | GAA ACG GAC AGT AAC CCG TAT GGC CGC CTA ATG TTG GGG GTG TTT GTC | 1056 |
|    | 337  | Glu Thr Asp Ser Asn Pro Tyr Gly Arg Leu Met Leu Gly Val Phe Val | 352  |
| 20 | 1057 | CGT CAA TTT GAA GGT GAT GGG GCC AAT AGA AAA AAT AAA CCC GTT GTT | 1104 |
|    | 353  | Arg Gln Phe Glu Gly Asp Gly Ala Asn Arg Lys Asn Lys Pro Val Val | 368  |
| 25 | 1105 | TAT GGT TAT CTC TAT TGT GAC TCA GCT TTC AAT CGT CAT GTT CTC AGG | 1152 |
|    | 369  | Tyr Gly Tyr Leu Tyr Cys Asp Ser Ala Phe Asn Arg His Val Leu Arg | 384  |
| 30 | 1153 | CCG TTA AGT AAG AAC TTT TTG TTC AGT ACT TAC CGT GAT GAA ACG GAT | 1200 |
|    | 385  | Pro Leu Ser Lys Asn Phe Leu Phe Ser Thr Tyr Arg Asp Glu Thr Asp | 400  |
| 35 | 1201 | GGT CAA AAC AGC TTG CAA TTT GCG GTA TAC GAT AAA AAG TAT GTA ATT | 1248 |
|    | 401  | Gly Gln Asn Ser Leu Gln Phe Ala Val Tyr Asp Lys Lys Tyr Val Ile | 416  |
| 40 | 1249 | ACT AAG GTT GTT ACA GGT GCA ACG GAA GAT CCC GAA AAT ACA GGA TGG | 1296 |
|    | 417  | Thr Lys Val Val Thr Gly Ala Thr Glu Asp Pro Glu Asn Thr Gly Trp | 432  |
| 45 | 1297 | GTA AGT AAA GTT GAT GAC TTG AAA CAA GGC ACT ACT GGG GCC TAT GTG | 1344 |
|    | 433  | Val Ser Lys Val Asp Asp Leu Lys Gln Gly Thr Thr Gly Ala Tyr Val | 448  |
| 50 | 1345 | TAT ATC GAT CAA GAT GGC CTG ACG CTT CAT ATA CAA ACC ACA ACT AAT | 1392 |
|    | 449  | Tyr Ile Asp Gln Asp Gly Leu Thr Leu His Ile Gln Thr Thr Thr Asn | 464  |
| 55 | 1393 | GGG GAT TTT ATT AAC CGT CAT ACG TTT GGA TAT AAC GAT CTT GTA TAT | 1440 |
|    | 465  | Gly Asp Phe Ile Asn Arg His Thr Phe Gly Tyr Asn Asp Leu Val Tyr | 480  |
| 60 | 1441 | GAT TCT AAG TCT GGT TAT GGT TTC ACG TGG TCA GGA AAT GAA GGT TTT | 1488 |
|    | 481  | Asp Ser Lys Ser Gly Tyr Gly Phe Thr Trp Ser Gly Asn Glu Gly Phe | 496  |
| 65 | 1489 | TAT CTG GAT TAC CAT GAT GGA AAT TAT TAC ACC TTT CAT AAT GCA ATA | 1536 |
|    | 497  | Tyr Leu Asp Tyr His Asp Gly Asn Tyr Tyr Thr Phe His Asn Ala Ile | 512  |
| 70 | 1537 | ATC AAC TAC TAT CCG TCT GGA TAT GGT GGT GGA TCT GTT CCT AAT GGA | 1584 |
|    | 513  | Ile Asn Tyr Tyr Pro Ser Gly Tyr Gly Gly Ser Val Pro Asn Gly     | 528  |
| 75 | 1585 | ACG TGG GCG TTA GAG CAA AGG ATT AAT GAG GGA TGG GCT ATT GCT CCC | 1632 |
|    | 529  | Thr Trp Ala Leu Glu Gln Arg Ile Asn Glu Gly Trp Ala Ile Ala Pro | 544  |
| 80 | 1633 | CTG CTT GAT ACT CTC CAT ACT GTT ACT GTG AAG GGC AGT TAT ATC GCT | 1680 |
|    | 545  | Leu Leu Asp Thr Leu His Thr Val Thr Val Lys Gly Ser Tyr Ile Ala | 560  |



|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
|    | 1631 | TCG GAA GGG GAA ACA CCT ACC GGT TAT AAT CTG TAT ATT CCA GAT GGT | 1728 |
|    | 561  | Trp Glu Gly Glu Thr Pro Thr Gly Tyr Asn Leu Tyr Ile Pro Asp Gly | 576  |
| 5  | 1729 | ACC GTG TTG CTA GAT TGG TTT GAT AAA ATA AAT TTT GCT ATT GGT CTT | 1776 |
|    | 577  | Thr Val Leu Leu Asp Trp Phe Asp Lys Ile Asn Phe Ala Ile Gly Leu | 592  |
| 10 | 1777 | AAT AAG CTT GAG TCT GTA TTT ACG TCG CCA GAT TGG CCA ACA CTA ACC | 1824 |
|    | 593  | Asn Lys Leu Glu Ser Val Phe Thr Ser Pro Asp Trp Pro Thr Leu Thr | 608  |
| 15 | 1825 | ACT ATC AAA AAT TTC AGT AAA ATC GCC GAT AAC CGC AAA TTC TAT CAG | 1872 |
|    | 609  | Thr Ile Lys Asn Phe Ser Lys Ile Ala Asp Asn Arg Lys Phe Tyr Gln | 624  |
|    | 1873 | GAA ATC AAT GCT GAG ACG GCG GAT GGA CGC AAC CTG TTT AAA CGT TAC | 1920 |
|    | 625  | Glu Ile Asn Ala Glu Thr Ala Asp Gly Arg Asn Leu Phe Lys Arg Tyr | 640  |
| 20 | 1921 | AGT ACT CAA ACT TTC GGA CTT ACC AGC GGT GCG ACT TAT TCT ACA ACT | 1968 |
|    | 641  | Ser Thr Gln Thr Phe Gly Leu Thr Ser Gly Ala Thr Tyr Ser Thr Thr | 656  |
| 25 | 1969 | TAT ACT TTG TCT GAG GCG GAT TTC TCC ACT GAT CCG GAC AAA AAC TAC | 2016 |
|    | 657  | Tyr Thr Leu Ser Glu Ala Asp Phe Ser Thr Asp Pro Asp Lys Asn Tyr | 672  |
| 30 | 2017 | CTA CAG GTT TGT TTG AAT GTC GTG TGG GAT CAT TAT GAC CGC CCG TCA | 2064 |
|    | 673  | Leu Gln Val Cys Leu Asn Val Val Trp Asp His Tyr Asp Arg Pro Ser | 688  |
| 35 | 2065 | GGG AAA AAA GGG GCT TAT TCT TGG GTC AGT AAG TGG TTT AAC GTC TAT | 2112 |
|    | 689  | Gly Lys Lys Gly Ala Tyr Ser Trp Val Ser Lys Trp Phe Asn Val Tyr | 704  |
|    | 2113 | GTT GCG TTG CAA GAT AGC AAA GCT CCG GAT GCC ATT CCT CGA TTA GTT | 2160 |
|    | 705  | Val Ala Leu Gln Asp Ser Lys Ala Pro Asp Ala Ile Pro Arg Leu Val | 720  |
| 40 | 2161 | TCC CGT TAC GAT AGT AAA CGT GGT CTG GTG CAA TAT CTG GAC TTC TGG | 2208 |
|    | 721  | Ser Arg Tyr Asp Ser Lys Arg Gly Leu Val Gln Tyr Leu Asp Phe Trp | 736  |
| 45 | 2209 | ACC TCA TCA TTA CCC GCG AAA ACC CGT CTT AAC ACC ACC TTT GTG CGT | 2256 |
|    | 737  | Thr Ser Ser Leu Pro Ala Lys Thr Arg Leu Asn Thr Thr Phe Val Arg | 752  |
| 50 | 2257 | ACT TTG ATT GAG AAG GCT AAT CTG GGG CTG GAT AGT TTG CTG GAT TAC | 2304 |
|    | 753  | Thr Leu Ile Glu Lys Ala Asn Leu Gly Leu Asp Ser Leu Leu Asp Tyr | 768  |
| 55 | 2305 | ACC TTG CAG GCA GAT CCT TCT CTG GAA GCA GAT TTA GTG ACT GAC GGC | 2352 |
|    | 769  | Thr Leu Gln Ala Asp Pro Ser Leu Glu Ala Asp Leu Val Thr Asp Gly | 784  |
|    | 2353 | AAA AGC GAA CCA ATG GAC TTT AAT GGT TCA AAC GGT CTC TAT TTC TGG | 2400 |
|    | 785  | Lys Ser Glu Pro Met Asp Phe Asn Gly Ser Asn Gly Leu Tyr Phe Trp | 800  |
| 60 | 2401 | GAA TTG TTC TTT CAC CTG CCG TTT TTG GTT GCT ACA CGC TTT GCC AAC | 2448 |
|    | 801  | Glu Leu Phe Phe His Leu Pro Phe Leu Val Ala Thr Arg Phe Ala Asn | 816  |
| 65 | 2449 | GAA CAG CAA TTT TCG CCG GCA CAA AAG AGT TTG CAT TAC ATC TTT GAC | 2496 |
|    | 817  | Glu Gln Gln Phe Ser Pro Ala Gln Lys Ser Leu His Tyr Ile Phe Asp | 832  |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
| 5  | 2497 | CCG GCG ATG AAA AAC AAG CCA CAC AAT GCC CCG GCT TAT TGG AAT GTA | 2544 |
|    | 333  | Pro Ala Met Lys Asn Lys Pro His Asn Ala Pro Ala Tyr Trp Asn Val | 348  |
| 10 | 2545 | CGT CCG TTG GTT GAA GGA AAC AGC GAT TTG TCA CGT CAT TTG GAC GAT | 2592 |
|    | 349  | Arg Pro Leu Val Glu Gly Asn Ser Asp Leu Ser Arg His Leu Asp Asp | 364  |
| 15 | 2593 | TCT ATA GAC CCA GAT ACT CAA GCT TAT GCT CAT CCG GTG ATA TAC CAG | 2640 |
|    | 365  | Ser Ile Asp Pro Asp Thr Gln Ala Tyr Ala His Pro Val Ile Tyr Gln | 380  |
| 20 | 2641 | AAA GCG GTG TTT ATT GCC TAT GTC AGT AAC CTG ATT GCT CAG GGA GAT | 2688 |
|    | 381  | Lys Ala Val Phe Ile Ala Tyr Val Ser Asn Leu Ile Ala Gln Gly Asp | 396  |
| 25 | 2689 | ATG TGG TAT CCG CAA TTG ACT CGT GAC GGT CTG ACT CAG GCC CGT GTC | 2736 |
|    | 397  | Met Trp Tyr Arg Gln Leu Thr Arg Asp Gly Leu Thr Gln Ala Arg Val | 312  |
| 30 | 2737 | TAT TAC AAT CTG GCC GCT GAA TTG CTA GGG CCT CGT CCG GAT GTA TCG | 2784 |
|    | 913  | Tyr Tyr Asn Leu Ala Ala Glu Leu Leu Gly Pro Arg Pro Asp Val Ser | 928  |
| 35 | 2785 | CTG AGT AGC ATT TGG ACG CCG CAA ACC CTG GAT ACC TTA GCA GCC GGG | 2832 |
|    | 929  | Leu Ser Ser Ile Trp Thr Pro Gln Thr Leu Asp Thr Leu Ala Ala Gly | 944  |
| 40 | 2833 | CAA AAA GCG GTT TTA CGT GAT TTT GAG CAC CAG TTG GCT AAT AGT GAT | 2880 |
|    | 945  | Gln Lys Ala Val Leu Arg Asp Phe Glu His Gln Leu Ala Asn Ser Asp | 960  |
| 45 | 2881 | ACC GCT TTA CCC GCA TTG CCG GGC CGC AAT GTC AGC TAC TTG AAA CTG | 2928 |
|    | 961  | Thr Ala Leu Pro Ala Leu Pro Gly Arg Asn Val Ser Tyr Leu Lys Leu | 976  |
| 50 | 2929 | GCA GAT AAT GGC TAC TTT AAT GAA CCG CTC AAT GTT CTG ATG TTG TCT | 2976 |
|    | 977  | Ala Asp Asn Gly Tyr Phe Asn Glu Pro Leu Asn Val Leu Met Leu Ser | 992  |
| 55 | 2977 | CAC TGG GAT ACG TTG GAT GCA CGG TTA TAC AAT CTG CGT CAT AAC CTG | 3024 |
|    | 993  | His Trp Asp Thr Leu Asp Ala Arg Leu Tyr Asn Leu Arg His Asn Leu | 1008 |
| 60 | 3025 | ACC GTT GAT GGC AAG CCG CTT TCG CTG CCG CTG TAT GCT GCG CCT GTT | 3072 |
|    | 1009 | Thr Val Asp Gly Lys Pro Leu Ser Leu Pro Leu Tyr Ala Ala Pro Val | 1024 |
| 65 | 3073 | GAT CCG GTA GCG TTG TTG GCT CAG CGT GCT CAG TCC GGC ACG TTG ACG | 3120 |
|    | 1025 | Asp Pro Val Ala Leu Leu Ala Gln Arg Ala Gln Ser Gly Thr Leu Thr | 1040 |
| 70 | 3121 | AAT GGC GTC AGT GGC GCC ATG TTG ACG GTG CCG CCA TAC CGT TTC AGC | 3168 |
|    | 1041 | Asn Gly Val Ser Gly Ala Met Leu Thr Val Pro Pro Tyr Arg Phe Ser | 1056 |
| 75 | 3169 | GCT ATG TTG CCG CGA GCT TAC AGC GCC GTG GGT ACG TTG ACC AGT TTT | 3216 |
|    | 1057 | Ala Met Leu Pro Arg Ala Tyr Ser Ala Val Gly Thr Leu Thr Ser Phe | 1072 |
| 80 | 3217 | GGT CAG AAC CTG CTT AGT TTG TTG GAA CGT AGC GAA CGA GCC TGT CAA | 3264 |
|    | 1073 | Gly Gln Asn Leu Leu Ser Leu Leu Glu Arg Ser Glu Arg Ala Cys Gln | 1088 |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
|    | 3265 | GAA GAG TTG GCG CAA CAG CAA CTG TTG GAT ATG TCC AGC TAT GCG ATC | 3311 |
|    | 1089 | Glu Glu Leu Ala Gln Gln Gln Leu Leu Asp Met Ser Ser Tyr Ala Ile | 1104 |
| 5  | 3313 | ACG TTG CAA CAA CAG GCG CTG GAT GGA TTG GCG GCA GAT CGT CTG GCG | 3360 |
|    | 1105 | Thr Leu Gln Gln Gln Ala Leu Asp Gly Leu Ala Ala Asp Arg Leu Ala | 1120 |
| 10 | 3361 | CTG CTA GCT AGT CAG GCT ACG GCA CAA CAG CGT CAT GAC CAT TAT TAC | 3408 |
|    | 1121 | Leu Leu Ala Ser Gln Ala Thr Ala Gln Gln Arg His Asp His Tyr Tyr | 1136 |
| 15 | 3409 | ACT CTG TAT CAG AAC AAC ATC TCC AGT GCG GAA CAA CTG GTG ATG GAC | 3456 |
|    | 1137 | Thr Leu Tyr Gln Asn Asn Ile Ser Ser Ala Glu Gln Leu Val Met Asp | 1152 |
|    | 3457 | ACC CAA ACG TCA GCA CAA TCC CTG ATT TCT TCT TCC ACT GGT GTA CAA | 3504 |
|    | 1153 | Thr Gln Thr Ser Ala Gln Ser Leu Ile Ser Ser Ser Thr Gly Val Gln | 1168 |
| 20 | 3505 | ACT GCC AGT GGG GCA CTG AAA GTG ATC CCG AAT ATC TTT GGT TTG GCT | 3552 |
|    | 1169 | Thr Ala Ser Gly Ala Leu Lys Val Ile Pro Asn Ile Phe Gly Leu Ala | 1184 |
| 25 | 3553 | GAT GGC GGC TCG CGC TAT GAA GGA GTA ACG GAA GCG ATT GCC ATC GGG | 3600 |
|    | 1185 | Asp Gly Gly Ser Arg Tyr Glu Gly Val Thr Glu Ala Ile Ala Ile Gly | 1200 |
| 30 | 3601 | TTA ATG GCT GCC GGA CAA GCC ACC AGC GTG GTG GCC GAG CGT CTG GCA | 3648 |
|    | 1201 | Leu Met Ala Ala Gly Gln Ala Thr Ser Val Val Ala Glu Arg Leu Ala | 1216 |
| 35 | 3649 | ACC ACG GAG AAT TAC CGC CGC CGC CGT GAA GAG TGG CAA ATC CAA TAC | 3696 |
|    | 1217 | Thr Thr Glu Asn Tyr Arg Arg Arg Arg Glu Glu Trp Gln Ile Gln Tyr | 1232 |
|    | 3697 | CAG CAG GCA CAG TCT GAG GTC GAC GCA TTA CAG AAA CAG TTG GAT GCG | 3744 |
|    | 1233 | Gln Gln Ala Gln Ser Glu Val Asp Ala Leu Gln Lys Gln Leu Asp Ala | 1248 |
| 40 | 3745 | CTG GCA GTG CGC GAG AAA GCA GCT CAA ACT TCC CTG CAA CAG GCG AAG | 3792 |
|    | 1249 | Leu Ala Val Arg Glu Lys Ala Ala Gln Thr Ser Leu Gln Gln Ala Lys | 1264 |
| 45 | 3793 | GCA CAG CAG GTA CAA ATT CGG ACC ATG CTG ACT TAC TTA ACT ACT CGT | 3840 |
|    | 1265 | Ala Gln Gln Val Gln Ile Arg Thr Met Leu Thr Tyr Leu Thr Thr Arg | 1280 |
| 50 | 3841 | TTC ACC CAG GCG ACT CTG TAC CAG TGG CTG AGT GGT CAA TTA TCC GCG | 3888 |
|    | 1281 | Phe Thr Gln Ala Thr Leu Tyr Gln Trp Leu Ser Gly Gln Leu Ser Ala | 1296 |
| 55 | 3889 | TTG TAT TAT CAA GCG TAT GAT GCC GTG GTT GCT CTC TGC CTC TCC GCC | 3936 |
|    | 1297 | Leu Tyr Tyr Gln Ala Tyr Asp Ala Val Val Ala Leu Cys Leu Ser Ala | 1312 |
|    | 3937 | CAA GCT TGC TGG CAG TAT GAA TTG GGT GAT TAC GCT ACC ACT TTT ATC | 3984 |
|    | 1313 | Gln Ala Cys Trp Gln Tyr Glu Leu Gly Asp Tyr Ala Thr Thr Phe Ile | 1323 |
| 60 | 3985 | CAG ACC GGT ACC TGG AAC GAC CAT TAC CGT GGT TTG CAA GTG GGG GAG | 4032 |
|    | 1329 | Gln Thr Gly Thr Trp Asn Asp His Tyr Arg Gly Leu Gln Val Gly Glu | 1344 |
| 65 | 4033 | ACA CTG CAA CTC AAT TTG CAT CAG ATG GAA GCG GCC TAT TTA GTT CGT | 4080 |
|    | 1345 | Thr Leu Gln Leu Asn Leu His Gln Met Glu Ala Ala Tyr Leu Val Arg | 1360 |

5  
 4081 CAC GAA CGC CGT CTT AAT GTG ATC CGT ACT GTG TCG CTC AAA AGC CTA 4128  
 1361 His Glu Arg Arg Leu Asn Val Ile Arg Thr Val Ser Leu Lys Ser Leu 1376

10  
 4129 TTG GGT GAT GAT GGT TTT GGT AAG TTA AAA ACC GAA GGC AAA GTC GAC 4176  
 1377 Leu Gly Asp Asp Gly Phe Gly Lys Leu Lys Thr Glu Gly Lys Val Asp 1392

15  
 4177 TTT CCA TTA AGC GAA AAG CTG TTT GAC AAC GAC TAT CCG GGG CAC TAT 4224  
 1393 Phe Pro Leu Ser Glu Lys Leu Phe Asp Asn Asp Tyr Pro Gly His Tyr 1408

20  
 4225 TTG CGC CAG ATT AAA ACT GTG TCA GTG ACG TTG CCG ACG TTA GTC GGG 4272  
 1409 Leu Arg Gln Ile Lys Thr Val Ser Val Thr Leu Pro Thr Leu Val Gly 1424

25  
 4273 CCG TAT CAA AAC GTG AAG GCA ACG CTC ACT CAG ACC AGC AGC AGT ATA 4320  
 1425 Pro Tyr Gln Asn Val Lys Ala Thr Leu Thr Gln Thr Ser Ser Ser Ile 1440

30  
 4321 TTG TTA GCA GCA GAT ATC AAT GGT GTT AAA CGT CTC AAT GAT CCG ACA 4368  
 1441 Leu Leu Ala Ala Asp Ile Asn Gly Val Lys Arg Leu Asn Asp Pro Thr 1456

35  
 4369 GGT AAA GAG GGT GAT GCG ACG CAT ATT GTC ACC AAT CTG CGT GCC AGC 4416  
 1457 Gly Lys Glu Gly Asp Ala Thr His Ile Val Thr Asn Leu Arg Ala Ser 1472

40  
 4417 CAG CAG GTG GCG CTC TCT TCT GGC ATT AAT GAT GCC GGT AGC TTT GAG 4464  
 1473 Gln Gln Val Ala Leu Ser Ser Gly Ile Asn Asp Ala Gly Ser Phe Glu 1488

45  
 4465 TTG CGT TTG GAA GAT GAG CGC TAT CTA TCA TTT GAG GGG ACT GGA GCT 4512  
 1489 Leu Arg Leu Glu Asp Glu Arg Tyr Leu Ser Phe Glu Gly Thr Gly Ala 1504

50  
 4513 GTT TCC AAA TGG ACT CTT AAC TTC CCG CGT TCT GTG GAT GAG CAT ATT 4560  
 1505 Val Ser Lys Trp Thr Leu Asn Phe Pro Arg Ser Val Asp Glu His Ile 1520

55  
 4561 GAC GAT AAG ACA TTG AAA GCG GAT GAG ATG CAG GCC GCA CTG TTG GCG 4608  
 1521 Asp Asp Lys Thr Leu Lys Ala Asp Glu Met Gln Ala Ala Leu Leu Ala 1536

60  
 4609 AAT ATG GAT GAT GTG CTG GTG CAG GTG CAT TAT ACC GCC TGC GAC GGC 4656  
 1537 Asn Met Asp Asp Val Leu Val Gln Val His Tyr Thr Ala Cys Asp Gly 1552

65  
 4657 GGC GCC AGT TTC GCA AAC CAG GTC AAG AAA ACA CTC TCT TAA 4698  
 1553 Gly Ala Ser Phe Ala Asn Gln Val Lys Lys Thr Leu Ser End 1566

55 (2) INFORMATION FOR SEQ ID NO:59  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1665 amino acids  
 (B) TYPE: amino acid  
 (C) TOPOLOGY: linear

60 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59 (TccB peptide)

65 Features From To Description

i 11 SEQ ID NO:7

|    |     |                                                                 |     |
|----|-----|-----------------------------------------------------------------|-----|
| 5  | 1   | Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu Ser Gln Arg Asp Ala | 16  |
|    | 17  | Leu Val Thr Gly Tyr Met Asn Phe Val Ala Pro Thr Leu Lys Gly Val | 32  |
|    | 33  | Ser Gly Gln Pro Val Thr Val Glu Asp Leu Tyr Glu Tyr Leu Leu Ile | 48  |
| 10 | 49  | Asp Pro Glu Val Ala Asp Glu Val Glu Thr Ser Arg Val Ala Gln Ala | 64  |
|    | 65  | Ile Ala Ser Ile Gln Gln Tyr Met Thr Arg Leu Val Asn Gly Ser Glu | 80  |
|    | 81  | Pro Gly Arg Gln Ala Met Glu Pro Ser Thr Ala Asn Glu Trp Arg Asp | 96  |
| 15 | 97  | Asn Asp Asn Gln Tyr Ala Ile Trp Ala Ala Gly Ala Glu Val Arg Asn | 112 |
|    | 113 | Tyr Ala Glu Asn Tyr Ile Ser Pro Ile Thr Arg Gln Glu Lys Ser His | 128 |
| 20 | 129 | Tyr Phe Ser Glu Leu Glu Thr Thr Leu Asn Gln Asn Arg Leu Asp Pro | 144 |
|    | 145 | Asp Arg Val Gln Asp Ala Val Leu Ala Tyr Leu Asn Glu Phe Glu Ala | 160 |
|    | 161 | Val Ser Asn Leu Tyr Val Leu Ser Gly Tyr Ile Asn Gln Asp Lys Phe | 176 |
| 25 | 177 | Asp Gln Ala Ile Tyr Tyr Phe Ile Gly Arg Thr Thr Thr Lys Pro Tyr | 192 |
|    | 193 | Arg Tyr Tyr Trp Arg Gln Met Asp Leu Ser Lys Asn Arg Gln Asp Pro | 208 |
| 30 | 209 | Ala Gly Asn Pro Val Thr Pro Asn Cys Trp Asn Asp Trp Gln Glu Ile | 224 |
|    | 225 | Thr Leu Pro Leu Ser Gly Asp Thr Val Leu Glu His Thr Val Arg Pro | 240 |
|    | 241 | Val Phe Tyr Asn Asp Arg Leu Tyr Val Ala Trp Val Glu Arg Asp Pro | 256 |
| 35 | 257 | Ala Val Gln Lys Asp Ala Asp Gly Lys Asn Ile Gly Lys Thr His Ala | 272 |
|    | 273 | Tyr Asn Ile Lys Phe Gly Tyr Lys Arg Tyr Asp Asp Thr Trp Thr Ala | 288 |
| 40 | 289 | Pro Asn Thr Thr Thr Leu Met Thr Gln Gln Ala Gly Glu Ser Ser Glu | 304 |
|    | 305 | Thr Gln Arg Ser Ser Leu Leu Ile Asp Glu Ser Ser Thr Thr Leu Arg | 320 |
|    | 321 | Gln Val Asn Leu Leu Ala Thr Thr Asp Phe Ser Ile Asp Pro Thr Glu | 336 |
| 45 | 337 | Glu Thr Asp Ser Asn Pro Tyr Gly Arg Leu Met Leu Gly Val Phe Val | 352 |
|    | 353 | Arg Gln Phe Glu Gly Asp Gly Ala Asn Arg Lys Asn Lys Pro Val Val | 368 |
| 50 | 369 | Tyr Gly Tyr Leu Tyr Cys Asp Ser Ala Phe Asn Arg His Val Leu Arg | 384 |
|    | 385 | Pro Leu Ser Lys Asn Phe Leu Phe Ser Thr Tyr Arg Asp Glu Thr Asp | 400 |
|    | 401 | Gly Gln Asn Ser Leu Gln Phe Ala Val Tyr Asp Lys Lys Tyr Val Ile | 416 |
| 55 | 417 | Thr Lys Val Val Thr Gly Ala Thr Glu Asp Pro Glu Asn Thr Gly Trp | 432 |
|    | 433 | Val Ser Lys Val Asp Asp Leu Lys Gln Gly Thr Thr Gly Ala Tyr Val | 448 |
| 60 | 449 | Tyr Ile Asp Gln Asp Gly Leu Thr Leu His Ile Gln Thr Thr Thr Asn | 464 |
|    | 465 | Gly Asp Phe Ile Asn Arg His Thr Phe Gly Tyr Asn Asp Leu Val Tyr | 480 |
|    | 481 | Asp Ser Lys Ser Gly Tyr Gly Phe Thr Trp Ser Gly Asn Glu Gly Phe | 496 |
| 65 | 497 | Tyr Leu Asp Tyr His Asp Gly Asn Tyr Tyr Thr Phe His Asn Ala Ile | 512 |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
|    | 513  | Ile Asn Tyr Tyr Pro Ser Gly Tyr Gly Gly Gly Ser Val Pro Asn Gly | 523  |
| 5  | 529  | Thr Trp Ala Leu Glu Gln Arg Ile Asn Glu Gly Trp Ala Ile Ala Pro | 544  |
|    | 545  | Leu Leu Asp Thr Leu His Thr Val Thr Val Lys Gly Ser Tyr Ile Ala | 560  |
|    | 561  | Trp Glu Gly Glu Thr Pro Thr Gly Tyr Asn Leu Tyr Ile Pro Asp Gly | 576  |
| 10 | 577  | Thr Val Leu Leu Asp Trp Phe Asp Lys Ile Asn Phe Ala Ile Gly Leu | 592  |
|    | 593  | Asn Lys Leu Glu Ser Val Phe Thr Ser Pro Asp Trp Pro Thr Leu Thr | 608  |
| 15 | 609  | Thr Ile Lys Asn Phe Ser Lys Ile Ala Asp Asn Arg Lys Phe Tyr Gln | 624  |
|    | 625  | Glu Ile Asn Ala Glu Thr Ala Asp Gly Arg Asn Leu Phe Lys Arg Tyr | 640  |
|    | 641  | Ser Thr Gln Thr Phe Gly Leu Thr Ser Gly Ala Thr Tyr Ser Thr Thr | 656  |
| 20 | 657  | Tyr Thr Leu Ser Glu Ala Asp Phe Ser Thr Asp Pro Asp Lys Asn Tyr | 672  |
|    | 673  | Leu Gln Val Cys Leu Asn Val Val Trp Asp His Tyr Asp Arg Pro Ser | 688  |
| 25 | 689  | Gly Lys Lys Gly Ala Tyr Ser Trp Val Ser Lys Trp Phe Asn Val Tyr | 704  |
|    | 705  | Val Ala Leu Gln Asp Ser Lys Ala Pro Asp Ala Ile Pro Arg Leu Val | 720  |
|    | 721  | Ser Arg Tyr Asp Ser Lys Arg Gly Leu Val Gln Tyr Leu Asp Phe Trp | 736  |
| 30 | 737  | Thr Ser Ser Leu Pro Ala Lys Thr Arg Leu Asn Thr Thr Phe Val Arg | 752  |
|    | 753  | Thr Leu Ile Glu Lys Ala Asn Leu Gly Leu Asp Ser Leu Leu Asp Tyr | 768  |
| 35 | 769  | Thr Leu Gln Ala Asp Pro Ser Leu Glu Ala Asp Leu Val Thr Asp Gly | 784  |
|    | 785  | Lys Ser Glu Pro Met Asp Phe Asn Gly Ser Asn Gly Leu Tyr Phe Trp | 800  |
|    | 801  | Glu Leu Phe Phe His Leu Pro Phe Leu Val Ala Thr Arg Phe Ala Asn | 816  |
| 40 | 817  | Glu Gln Gln Phe Ser Pro Ala Gln Lys Ser Leu His Tyr Ile Phe Asp | 832  |
|    | 833  | Pro Ala Met Lys Asn Lys Pro His Asn Ala Pro Ala Tyr Trp Asn Val | 848  |
| 45 | 849  | Arg Pro Leu Val Glu Gly Asn Ser Asp Leu Ser Arg His Leu Asp Asp | 864  |
|    | 865  | Ser Ile Asp Pro Asp Thr Gln Ala Tyr Ala His Pro Val Ile Tyr Gln | 880  |
|    | 881  | Lys Ala Val Phe Ile Ala Tyr Val Ser Asn Leu Ile Ala Gln Gly Asp | 896  |
| 50 | 897  | Met Trp Tyr Arg Gln Leu Thr Arg Asp Gly Leu Thr Gln Ala Arg Val | 912  |
|    | 913  | Tyr Tyr Asn Leu Ala Ala Glu Leu Leu Gly Pro Arg Pro Asp Val Ser | 928  |
| 55 | 929  | Leu Ser Ser Ile Trp Thr Pro Gln Thr Leu Asp Thr Leu Ala Ala Gly | 944  |
|    | 945  | Gln Lys Ala Val Leu Arg Asp Phe Glu His Gln Leu Ala Asn Ser Asp | 960  |
|    | 961  | Thr Ala Leu Pro Ala Leu Pro Gly Arg Asn Val Ser Tyr Leu Lys Leu | 976  |
| 60 | 977  | Ala Asp Asn Gly Tyr Phe Asn Glu Pro Leu Asn Val Leu Met Leu Ser | 992  |
|    | 993  | His Trp Asp Thr Leu Asp Ala Arg Leu Tyr Asn Leu Arg His Asn Leu | 1008 |
| 65 | 1009 | Thr Val Asp Gly Lys Pro Leu Ser Leu Pro Leu Tyr Ala Ala Pro Val | 1024 |
|    | 1025 | Asp Pro Val Ala Leu Leu Ala Gln Arg Ala Gln Ser Gly Thr Leu Thr | 1040 |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
|    | 1041 | Asn Gly Val Ser Gly Ala Met Leu Thr Val Pro Pro Tyr Arg Phe Ser | 1056 |
|    | 1057 | Ala Met Leu Pro Arg Ala Tyr Ser Ala Val Gly Thr Leu Thr Ser Phe | 1072 |
| 5  | 1073 | Gly Gln Asn Leu Leu Ser Leu Leu Glu Arg Ser Glu Arg Ala Cys Gln | 1088 |
|    | 1089 | Glu Glu Leu Ala Gln Gln Gln Leu Leu Asp Met Ser Ser Tyr Ala Ile | 1104 |
| 10 | 1105 | Thr Leu Gln Gln Gln Ala Leu Asp Gly Leu Ala Ala Asp Arg Leu Ala | 1120 |
|    | 1121 | Leu Leu Ala Ser Gln Ala Thr Ala Gln Gln Arg His Asp His Tyr Tyr | 1136 |
|    | 1137 | Thr Leu Tyr Gln Asn Asn Ile Ser Ser Ala Glu Gln Leu Val Met Asp | 1152 |
| 15 | 1153 | Thr Gln Thr Ser Ala Gln Ser Leu Ile Ser Ser Ser Thr Gly Val Gln | 1168 |
|    | 1169 | Thr Ala Ser Gly Ala Leu Lys Val Ile Pro Asn Ile Phe Gly Leu Ala | 1184 |
| 20 | 1185 | Asp Gly Gly Ser Arg Tyr Glu Gly Val Thr Glu Ala Ile Ala Ile Gly | 1200 |
|    | 1201 | Leu Met Ala Ala Gly Gln Ala Thr Ser Val Val Ala Glu Arg Leu Ala | 1216 |
|    | 1217 | Thr Thr Glu Asn Tyr Arg Arg Arg Arg Glu Glu Trp Gln Ile Gln Tyr | 1232 |
| 25 | 1233 | Gln Gln Ala Gln Ser Glu Val Asp Ala Leu Gln Lys Gln Leu Asp Ala | 1248 |
|    | 1249 | Leu Ala Val Arg Glu Lys Ala Ala Gln Thr Ser Leu Gln Gln Ala Lys | 1264 |
| 30 | 1265 | Ala Gln Gln Val Gln Ile Arg Thr Met Leu Thr Tyr Leu Thr Thr Arg | 1280 |
|    | 1281 | Phe Thr Gln Ala Thr Leu Tyr Gln Trp Leu Ser Gly Gln Leu Ser Ala | 1296 |
|    | 1297 | Leu Tyr Tyr Gln Ala Tyr Asp Ala Val Val Ala Leu Cys Leu Ser Ala | 1312 |
| 35 | 1313 | Gln Ala Cys Trp Gln Tyr Glu Leu Gly Asp Tyr Ala Thr Thr Phe Ile | 1328 |
|    | 1329 | Gln Thr Gly Thr Trp Asn Asp His Tyr Arg Gly Leu Gln Val Gly Glu | 1344 |
| 40 | 1345 | Thr Leu Gln Leu Asn Leu His Gln Met Glu Ala Ala Tyr Leu Val Arg | 1360 |
|    | 1361 | His Glu Arg Arg Leu Asn Val Ile Arg Thr Val Ser Leu Lys Ser Leu | 1376 |
|    | 1377 | Leu Gly Asp Asp Gly Phe Gly Lys Leu Lys Thr Glu Gly Lys Val Asp | 1392 |
| 45 | 1393 | Phe Pro Leu Ser Glu Lys Leu Phe Asp Asn Asp Tyr Pro Gly His Tyr | 1408 |
|    | 1409 | Leu Arg Gln Ile Lys Thr Val Ser Val Thr Leu Pro Thr Leu Val Gly | 1424 |
| 50 | 1425 | Pro Tyr Gln Asn Val Lys Ala Thr Leu Thr Gln Thr Ser Ser Ser Ile | 1440 |
|    | 1441 | Leu Leu Ala Ala Asp Ile Asn Gly Val Lys Arg Leu Asn Asp Pro Thr | 1456 |
|    | 1457 | Gly Lys Glu Gly Asp Ala Thr His Ile Val Thr Asn Leu Arg Ala Ser | 1472 |
| 55 | 1473 | Gln Gln Val Ala Leu Ser Ser Gly Ile Asn Asp Ala Gly Ser Phe Glu | 1488 |
|    | 1489 | Leu Arg Leu Glu Asp Glu Arg Tyr Leu Ser Phe Glu Gly Thr Gly Ala | 1504 |
| 60 | 1505 | Val Ser Lys Trp Thr Leu Asn Phe Pro Arg Ser Val Asp Glu His Ile | 1520 |
|    | 1521 | Asp Asp Lys Thr Leu Lys Ala Asp Glu Met Gln Ala Ala Leu Leu Ala | 1536 |
|    | 1537 | Asn Met Asp Asp Val Leu Val Gln Val His Tyr Thr Ala Cys Asp Gly | 1552 |
| 65 | 1553 | Gly Ala Ser Phe Ala Asn Gln Val Lys Lys Thr Leu Ser             | 1565 |

## (2) INFORMATION FOR SEQ ID NO:60

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 3132 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60 (cacc)

15 1 ATG AGT CCG TCT GAG ACT ACT CTT TAT ACT CAA ACC CCA ACA GTC AGC 48  
 1 Met Ser Pro Ser Glu Thr Thr Leu Tyr Thr Gln Thr Pro Thr Val Ser 16

20 49 GTG TTA GAT AAT CGC GGT CTG TCC ATT CGT GAT ATT GGT TTT CAC CGT 96  
 17 Val Leu Asp Asn Arg Gly Leu Ser Ile Arg Asp Ile Gly Phe His Arg 32

25 97 ATT GTA ATC GGG GGG GAT ACT GAC ACC CGC GTC ACC CGT CAC CAG TAT 144  
 33 Ile Val Ile Gly Gly Asp Thr Asp Thr Arg Val Thr Arg His Gln Tyr 48

30 145 GAT GCC CGT GGA CAC CTG AAC TAC AGT ATT GAC CCA CGC TTG TAT GAT 192  
 49 Asp Ala Arg Gly His Leu Asn Tyr Ser Ile Asp Pro Arg Leu Tyr Asp 64

35 193 GCA AAG CAG GCT GAT AAC TCA GTA AAG CCT AAT TTT GTC TGG CAG CAT 240  
 65 Ala Lys Gln Ala Asp Asn Ser Val Lys Pro Asn Phe Val Trp Gln His 80

40 241 GAT CTG GCC GGT CAT GCC CTG CGG ACA GAG AGT GTC GAT GCT GGT CGT 288  
 81 Asp Leu Ala Gly His Ala Leu Arg Thr Glu Ser Val Asp Ala Gly Arg 96

45 289 ACT GTT GCA TTG AAT GAT ATT GAA GGT CGT TCG GTA ATG ACA ATG AAT 336  
 97 Thr Val Ala Leu Asn Asp Ile Glu Gly Arg Ser Val Met Thr Met Asn 112

337 GCG ACC GGT GTT CGT CAG ACC CGT CGC TAT GAA GGC AAC ACC TTG CCC 384  
 113 Ala Thr Gly Val Arg Gln Thr Arg Arg Tyr Glu Gly Asn Thr Leu Pro 123

50 385 GGT CGC TTG TTA TCT GTG AGC GAG CAA GTT TTC AAC CAA GAG AGT GCT 432  
 129 Gly Arg Leu Leu Ser Val Ser Glu Gln Val Phe Asn Gln Glu Ser Ala 144

55 433 AAA GTG ACA GAG CGC TTT ATC TGG GCT GGG AAT ACA ACC TCG GAG AAA 480  
 145 Lys Val Thr Glu Arg Phe Ile Trp Ala Gly Asn Thr Thr Ser Glu Lys 160

60 481 GAG TAT AAC CTC TCC GGT CTG TGT ATA CGC CAC TAC GAC ACA GCG GGA 528  
 161 Glu Tyr Asn Leu Ser Gly Leu Cys Ile Arg His Tyr Asp Thr Ala Gly 176

529 GTG ACC CGG TTG ATG AGT CAG TCA CTG GCG GGC GCC ATG CTA TCC CAA 576  
 177 Val Thr Arg Leu Met Ser Gln Ser Leu Ala Gly Ala Met Leu Ser Gln 192

65 577 TCT CAC CAA TTG CTG GCG GAA GGG CAG GAG GCT AAC TGG AGC GGT GAC 624  
 193 Ser His Gln Leu Leu Ala Glu Gly Gln Glu Ala Asn Trp Ser Gly Asp 208



|    |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | 625  | GAC | GAA | ACT | GTC | TGG | CAG | GGA | ATG | CTG | GCA | AGT | GAG | GTC | TAT | ACG | ACA | 672  |
|    | 209  | Asp | Glu | Thr | Val | Trp | Gln | Gly | Met | Leu | Ala | Ser | Glu | Val | Tyr | Thr | Thr | 224  |
| 5  | 673  | CAA | AGT | ACC | ACT | AAT | GCC | ATC | GGG | GCT | TTA | CTG | ACC | CAA | ACC | GAT | GCG | 720  |
|    | 225  | Gln | Ser | Thr | Thr | Asn | Ala | Ile | Gly | Ala | Leu | Leu | Thr | Gln | Thr | Asp | Ala | 240  |
| 10 | 721  | AAA | GGC | AAT | ATT | CAG | CGT | CTG | GCT | TAT | GAC | ATT | GCC | GGT | CAG | TTA | AAA | 768  |
|    | 241  | Lys | Gly | Asn | Ile | Gln | Arg | Leu | Ala | Tyr | Asp | Ile | Ala | Gly | Gln | Leu | Lys | 256  |
| 15 | 769  | GGG | AGT | TGG | TTG | ACG | GTG | AAA | GGC | CAG | AGT | GAA | CAG | GTG | ATT | GTT | AAG | 816  |
|    | 257  | Gly | Ser | Trp | Leu | Thr | Val | Lys | Gly | Gln | Ser | Glu | Gln | Val | Ile | Val | Lys | 272  |
| 20 | 817  | TCC | CTG | AGC | TGG | TCA | GCC | GCA | GGT | CAT | AAA | TTG | CGT | GAA | GAG | CAC | GGT | 864  |
|    | 273  | Ser | Leu | Ser | Trp | Ser | Ala | Ala | Gly | His | Lys | Leu | Arg | Glu | Glu | His | Gly | 288  |
| 25 | 865  | AAC | GGC | GTG | GTT | ACG | GAG | TAC | AGT | TAT | GAG | CCG | GAA | ACT | CAA | CGT | CTG | 912  |
|    | 289  | Asn | Gly | Val | Val | Thr | Glu | Tyr | Ser | Tyr | Glu | Pro | Glu | Thr | Gln | Arg | Leu | 304  |
| 30 | 913  | ATA | GGT | ATC | ACC | ACC | CGG | CGT | GCC | GAA | GGG | AGT | CAA | TCA | GGA | GCC | AGA | 960  |
|    | 305  | Ile | Gly | Ile | Thr | Thr | Arg | Arg | Ala | Glu | Gly | Ser | Gln | Ser | Gly | Ala | Arg | 320  |
| 35 | 961  | GTA | TTG | CAG | GAT | CTA | CGC | TAT | AAG | TAT | GAT | CCG | GTG | GGG | AAT | GTT | ATC | 1008 |
|    | 321  | Val | Leu | Gln | Asp | Leu | Arg | Tyr | Lys | Tyr | Asp | Pro | Val | Gly | Asn | Val | Ile | 336  |
| 40 | 1009 | AGT | ATC | CAT | AAT | GAT | GCC | GAA | GCT | ACC | CGC | TTT | TGG | CGT | AAT | CAG | AAA | 1056 |
|    | 337  | Ser | Ile | His | Asn | Asp | Ala | Glu | Ala | Thr | Arg | Phe | Trp | Arg | Asn | Gln | Lys | 352  |
| 45 | 1057 | GTG | GAG | CCG | GAG | AAT | CGC | TAT | GTT | TAT | GAT | TCT | CTG | TAT | CAG | CTT | ATG | 1104 |
|    | 353  | Val | Glu | Pro | Glu | Asn | Arg | Tyr | Val | Tyr | Asp | Ser | Leu | Tyr | Gln | Leu | Met | 368  |
| 50 | 1105 | AGT | GCG | ACA | GGG | CGT | GAA | ATG | GCT | AAT | ATC | GGT | CAG | CAA | AGC | AAC | CAA | 1152 |
|    | 369  | Ser | Ala | Thr | Gly | Arg | Glu | Met | Ala | Asn | Ile | Gly | Gln | Gln | Ser | Asn | Gln | 384  |
| 55 | 1153 | CTT | CCC | TCA | CCC | GTT | ATA | CCT | GTT | CCT | ACT | GAC | GAC | AGC | ACT | TAT | ACC | 1200 |
|    | 385  | Leu | Pro | Ser | Pro | Val | Ile | Pro | Val | Pro | Thr | Asp | Asp | Ser | Thr | Tyr | Thr | 400  |
| 60 | 1201 | AAT | TAC | CTT | CGT | ACC | TAT | ACT | TAT | GAC | CGT | GGC | GGT | AAT | TTG | GTT | CAA | 1248 |
|    | 401  | Asn | Tyr | Leu | Arg | Thr | Tyr | Thr | Tyr | Asp | Arg | Gly | Gly | Asn | Leu | Val | Gln | 416  |
| 65 | 1249 | ATC | CGA | CAC | AGT | TCA | CCC | GCG | ACT | CAA | AAT | AGT | TAC | ACC | ACA | GAT | ATC | 1296 |
|    | 417  | Ile | Arg | His | Ser | Ser | Pro | Ala | Thr | Gln | Asn | Ser | Tyr | Thr | Thr | Asp | Ile | 432  |
| 70 | 1297 | ACC | GTT | TCA | AGC | CGC | AGT | AAC | CGG | GCG | GTA | TTG | AGT | ACA | TTA | ACG | ACA | 1344 |
|    | 433  | Thr | Val | Ser | Ser | Arg | Ser | Asn | Arg | Ala | Val | Leu | Ser | Thr | Leu | Thr | Thr | 448  |
| 75 | 1345 | GAT | CCA | ACC | CGA | GTG | GAT | GCG | CTA | TTT | GAT | TCC | GGC | GGT | CAT | CAG | AAG | 1392 |
|    | 449  | Asp | Pro | Thr | Arg | Val | Asp | Ala | Leu | Phe | Asp | Ser | Gly | Gly | His | Gln | Lys | 464  |
| 80 | 1393 | ATG | TTA | ATA | CCG | GGG | CAA | AAT | CTG | GAT | TGG | AAT | ATT | CGG | GGT | GAA | TTG | 1440 |

|    |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|    | 455  | Met | Leu | Ile | Pro | Gly | Gln | Asn | Leu | Asp | Trp | Asn | Ile | Arg | Gly | Glu | Leu | 460  |
| 5  | 1441 | CAA | CGA | GTC | ACA | CCG | GTG | AGC | CGT | GAA | AAT | AGC | AGT | GAC | AGT | GAA | TGG | 1438 |
|    | 431  | Gln | Arg | Val | Thr | Pro | Val | Ser | Arg | Glu | Asn | Ser | Ser | Asp | Ser | Glu | Trp | 496  |
| 10 | 1489 | TAT | CGC | TAT | AGC | AGT | GAT | GGC | ATG | CGG | CTG | CTA | AAA | GTG | AGT | GAA | CAG | 1536 |
|    | 497  | Tyr | Arg | Tyr | Ser | Ser | Asp | Gly | Met | Arg | Leu | Leu | Lys | Val | Ser | Glu | Gln | 512  |
| 15 | 1537 | CAG | ACG | GGC | AAC | AGT | ACT | CAA | GTA | CAA | CGG | GTG | ACT | TAT | CTG | CCG | GGA | 1584 |
|    | 513  | Gln | Thr | Gly | Asn | Ser | Thr | Gln | Val | Gln | Arg | Val | Thr | Tyr | Leu | Pro | Gly | 528  |
|    | 1585 | TTA | GAG | CTA | CGG | ACA | ACT | GGG | GTT | GCA | GAT | AAA | ACA | ACC | GAA | GAT | TTG | 1632 |
|    | 529  | Leu | Glu | Leu | Arg | Thr | Thr | Gly | Val | Ala | Asp | Lys | Thr | Thr | Glu | Asp | Leu | 544  |
| 20 | 1633 | CAG | GTG | ATT | ACG | GTA | GGT | GAA | GCG | GGT | CGC | GCA | CAG | GTA | AGG | GTA | TTG | 1680 |
|    | 545  | Gln | Val | Ile | Thr | Val | Gly | Glu | Ala | Gly | Arg | Ala | Gln | Val | Arg | Val | Leu | 560  |
| 25 | 1681 | CAC | TGG | GAA | AGT | GGT | AAG | CCG | ACA | GAT | ATT | GAC | AAC | AAT | CAG | GTG | CGC | 1728 |
|    | 561  | His | Trp | Glu | Ser | Gly | Lys | Pro | Thr | Asp | Ile | Asp | Asn | Asn | Gln | Val | Arg | 576  |
| 30 | 1729 | TAC | AGC | TAC | GAT | AAT | CTG | CTT | GGC | TCC | AGC | CAG | CTT | GAA | CTG | GAT | AGC | 1776 |
|    | 577  | Tyr | Ser | Tyr | Asp | Asn | Leu | Leu | Gly | Ser | Ser | Gln | Leu | Glu | Leu | Asp | Ser | 592  |
| 35 | 1777 | GAA | GGG | CAG | ATT | CTC | AGT | CAG | GAA | GAG | TAT | TAT | CCG | TAT | GGC | GGT | ACG | 1824 |
|    | 593  | Glu | Gly | Gln | Ile | Leu | Ser | Gln | Glu | Glu | Tyr | Tyr | Pro | Tyr | Gly | Gly | Thr | 608  |
|    | 1825 | GCG | ATA | TGG | GCG | GCG | AGA | AAT | CAG | ACA | GAA | GCC | AGC | TAC | AAA | TTT | ATT | 1872 |
|    | 609  | Ala | Ile | Trp | Ala | Ala | Arg | Asn | Gln | Thr | Glu | Ala | Ser | Tyr | Lys | Phe | Ile | 624  |
| 40 | 1873 | CGT | TAC | TCC | GGT | AAA | GAG | CGG | GAT | GCC | ACT | GGA | TTG | TAT | TAT | TAC | GGC | 1920 |
|    | 625  | Arg | Tyr | Ser | Gly | Lys | Glu | Arg | Asp | Ala | Thr | Gly | Leu | Tyr | Tyr | Tyr | Gly | 640  |
| 45 | 1921 | TAC | CGT | TAT | TAT | CAA | CCT | TGG | GTG | GGT | CGA | TGG | TTG | AGT | GCT | GAT | CCG | 1968 |
|    | 641  | Tyr | Arg | Tyr | Tyr | Gln | Pro | Trp | Val | Gly | Arg | Trp | Leu | Ser | Ala | Asp | Pro | 656  |
| 50 | 1969 | GCG | GGA | ACC | GTG | GAT | GGG | CTG | AAT | TTG | TAC | CGA | ATG | GTG | AGG | AAT | AAC | 2016 |
|    | 657  | Ala | Gly | Thr | Val | Asp | Gly | Leu | Asn | Leu | Tyr | Arg | Met | Val | Arg | Asn | Asn | 672  |
| 55 | 2017 | CCC | ATC | ACA | TTG | ACT | GAC | CAT | GAC | GGA | TTA | GCA | CCG | TCT | CCA | AAT | AGA | 2064 |
|    | 673  | Pro | Ile | Thr | Leu | Thr | Asp | His | Asp | Gly | Leu | Ala | Pro | Ser | Pro | Asn | Arg | 688  |
|    | 2065 | AAT | CGA | AAT | ACA | TTT | TGG | TTT | GCT | TCA | TTT | TTG | TTT | CGT | AAA | CCT | GAT | 2112 |
|    | 689  | Asn | Arg | Asn | Thr | Phe | Trp | Phe | Ala | Ser | Phe | Leu | Phe | Arg | Lys | Pro | Asp | 704  |
| 60 | 2113 | GAG | GGA | ATG | TCC | GCG | TCA | ATG | AGA | CGG | GGA | CAA | AAA | ATT | GGC | AGA | GCC | 2160 |
|    | 705  | Glu | Gly | Met | Ser | Ala | Ser | Met | Arg | Arg | Gly | Gln | Lys | Ile | Gly | Arg | Ala | 720  |
| 65 | 2161 | ATT | GCC | GGC | GGG | ATT | GCG | ATT | GGC | GGT | CTT | GCG | GCT | ACC | ATT | GCC | GCT | 2208 |
|    | 721  | Ile | Ala | Gly | Gly | Ile | Ala | Ile | Gly | Gly | Leu | Ala | Ala | Thr | Ile | Ala | Ala | 736  |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
| 5  | 2209 | ACG GCT GGC GCG GCT ATC CCC GTC ATT CTG GGG GTT GCG GCC STA GGC | 2256 |
|    | 757  | Thr Ala Gly Ala Ala Ile Pro Val Ile Leu Gly Val Ala Ala Val Gly | 752  |
| 10 | 2257 | GCG GGG ATT GGC GCG TTG ATG GGA TAT AAC GTC GGT AGC CTG CTG GAA | 2304 |
|    | 753  | Ala Gly Ile Gly Ala Leu Met Gly Tyr Asn Val Gly Ser Leu Leu Glu | 763  |
| 15 | 2305 | AAA GGC GGG GCA TTA CTT GCT CGA CTC GTA CAG GGG AAA TCG ACG TTA | 2352 |
|    | 769  | Lys Gly Gly Ala Leu Leu Ala Arg Leu Val Gln Gly Lys Ser Thr Leu | 784  |
| 20 | 2353 | GTA CAG TCG GCG GCT GGC GCG GCT GCC GGA GCG AGT TCA GCC GCG GCT | 2400 |
|    | 785  | Val Gln Ser Ala Ala Gly Ala Ala Ala Gly Ala Ser Ser Ala Ala Ala | 800  |
| 25 | 2401 | TAT GGC GCA CGG GCA CAA GGT GTC GGT GTT GCA TCA GCC GCC GGG GCG | 2448 |
|    | 801  | Tyr Gly Ala Arg Ala Gln Gly Val Gly Val Ala Ser Ala Ala Gly Ala | 816  |
| 30 | 2449 | GTA ACA GGG GCT GTG GGA TCA TGG ATA AAT AAT GCT GAT CGG GGG ATT | 2496 |
|    | 817  | Val Thr Gly Ala Val Gly Ser Trp Ile Asn Asn Ala Asp Arg Gly Ile | 832  |
| 35 | 2497 | GGC GGC GCT ATT GGG GCC GGG AGT GCG GTA GGC ACC ATT GAT ACT ATG | 2544 |
|    | 833  | Gly Gly Ala Ile Gly Ala Gly Ser Ala Val Gly Thr Ile Asp Thr Met | 848  |
| 40 | 2545 | TTA GGG ACT GCC TCT ACC CTT ACC CAT GAA GTC GGG GCA GCG GCG GGT | 2592 |
|    | 849  | Leu Gly Thr Ala Ser Thr Leu Thr His Glu Val Gly Ala Ala Ala Gly | 864  |
| 45 | 2593 | GGG GCG GCG GGT GGG ATG ATC ACC GGT ACG CAA GGG AGT ACT CGG GCA | 2640 |
|    | 865  | Gly Ala Ala Gly Gly Met Ile Thr Gly Thr Gln Gly Ser Thr Arg Ala | 880  |
| 50 | 2641 | GGT ATC CAT GCC GGT ATT GGC ACC TAT TAT GGC TCC TGG ATT GGT TTT | 2688 |
|    | 881  | Gly Ile His Ala Gly Ile Gly Thr Tyr Tyr Gly Ser Trp Ile Gly Phe | 896  |
| 55 | 2689 | GGT TTA GAT GTC GCT AGT AAC CCC GCC GGA CAT TTA GCG AAT TAC GCA | 2736 |
|    | 897  | Gly Leu Asp Val Ala Ser Asn Pro Ala Gly His Leu Ala Asn Tyr Ala | 912  |
| 60 | 2737 | GTG GGT TAT GCC GCT GGT TTG GGT GCT GAA ATG GCT GTC AAC AGA ATA | 2784 |
|    | 913  | Val Gly Tyr Ala Ala Gly Leu Gly Ala Glu Met Ala Val Asn Arg Ile | 928  |
| 65 | 2785 | ATG GGT GGT GGA TTT TTG AGT AGG CTC TTA GGC CGG GTT GTC AGC CCA | 2832 |
|    | 929  | Met Gly Gly Gly Phe Leu Ser Arg Leu Leu Gly Arg Val Val Ser Pro | 944  |
| 70 | 2833 | TAT GCC GCC GGT TTA GCC AGA CAA TTA GTA CAT TTC AGT GTC GCC AGA | 2880 |
|    | 945  | Tyr Ala Ala Gly Leu Ala Arg Gln Leu Val His Phe Ser Val Ala Arg | 960  |
| 75 | 2881 | CCT GTC TTT GAG CCG ATA TTT AGT GTT CTC GGC GGG CTT GTC GGT GGT | 2928 |
|    | 961  | Pro Val Phe Glu Pro Ile Phe Ser Val Leu Gly Gly Leu Val Gly Gly | 976  |
| 80 | 2929 | ATT GGA ACT GGC CTG CAC AGA GTG ATG GGA AGA GAG AGT TGG ATT TCC | 2976 |
|    | 977  | Ile Gly Thr Gly Leu His Arg Val Met Gly Arg Glu Ser Trp Ile Ser | 992  |
| 85 | 2977 | AGA GCG TTA AGT GCT GCC GGT AGT GGT ATA GAT CAT GTC GCT GGC ATG | 3024 |
|    |      |                                                                 |      |

393 Arg Ala Leu Ser Ala Ala Gly Ser Gly Ile Asp His Val Ala Gly Met 1013  
 5 3025 ATT GGT AAT CAG ATC AGA GGC AGG GTC TTG ACC ACA ACC GGG ATC GCT 3071  
 1009 Ile Gly Asn Gln Ile Arg Gly Arg Val Leu Thr Thr Thr Gly Ile Ala 1014  
 3073 AAT GCG ATA GAC TAT GGC ACC AGT GCT GTG GGA GCC GCA CGA CGA GTT 3120  
 10 1025 Asn Ala Ile Asp Tyr Gly Thr Ser Ala Val Gly Ala Ala Arg Arg Val 1040  
 3121 TTT TCT TTG TAA 3132  
 1041 Phe Ser Leu End 1043

15

## (2) INFORMATION FOR SEQ ID NO:61

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1043 amino acids  
 (B) TYPE: amino acid  
 20 (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61 (Tccc peptide)

1 Met Ser Pro Ser Glu Thr Thr Leu Tyr Thr Gln Thr Pro Thr Val Ser 16  
 17 Val Leu Asp Asn Arg Gly Leu Ser Ile Arg Asp Ile Gly Phe His Arg 32  
 30 33 Ile Val Ile Gly Gly Asp Thr Asp Thr Arg Val Thr Arg His Gln Tyr 48  
 49 Asp Ala Arg Gly His Leu Asn Tyr Ser Ile Asp Pro Arg Leu Tyr Asp 64  
 35 65 Ala Lys Gln Ala Asp Asn Ser Val Lys Pro Asn Phe Val Trp Gln His 80  
 81 Asp Leu Ala Gly His Ala Leu Arg Thr Glu Ser Val Asp Ala Gly Arg 96  
 40 97 Thr Val Ala Leu Asn Asp Ile Glu Gly Arg Ser Val Met Thr Met Asn 112  
 113 Ala Thr Gly Val Arg Gln Thr Arg Arg Tyr Glu Gly Asn Thr Leu Pro 128  
 129 Gly Arg Leu Leu Ser Val Ser Glu Gln Val Phe Asn Gln Glu Ser Ala 144  
 45 145 Lys Val Thr Glu Arg Phe Ile Trp Ala Gly Asn Thr Thr Ser Glu Lys 160  
 161 Glu Tyr Asn Leu Ser Gly Leu Cys Ile Arg His Tyr Asp Thr Ala Gly 176  
 177 Val Thr Arg Leu Met Ser Gln Ser Leu Ala Gly Ala Met Leu Ser Gln 192  
 50 193 Ser His Gln Leu Leu Ala Glu Gly Gln Glu Ala Asn Trp Ser Gly Asp 208  
 209 Asp Glu Thr Val Trp Gln Gly Met Leu Ala Ser Glu Val Tyr Thr Thr 224  
 55 225 Gln Ser Thr Thr Asn Ala Ile Gly Ala Leu Leu Thr Gln Thr Asp Ala 240  
 241 Lys Gly Asn Ile Gln Arg Leu Ala Tyr Asp Ile Ala Gly Gln Leu Lys 256  
 257 Gly Ser Trp Leu Thr Val Lys Gly Gln Ser Glu Gln Val Ile Val Lys 272  
 60 273 Ser Leu Ser Trp Ser Ala Ala Gly His Lys Leu Arg Glu Glu His Gly 288  
 289 Asn Gly Val Val Thr Glu Tyr Ser Tyr Glu Pro Glu Thr Gln Arg Leu 304  
 65 305 Ile Gly Ile Thr Thr Arg Arg Ala Glu Gly Ser Gln Ser Gly Ala Arg 320

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | 321 | Val | Leu | Gln | Asp | Leu | Arg | Tyr | Lys | Tyr | Asp | Pro | Val | Gly | Asn | Val | Ile | 336 |
| 5  | 337 | Ser | Ile | His | Asn | Asp | Ala | Glu | Ala | Thr | Arg | Phe | Trp | Arg | Asn | Gln | Lys | 352 |
|    | 353 | Val | Glu | Pro | Glu | Asn | Arg | Tyr | Val | Tyr | Asp | Ser | Leu | Tyr | Gln | Leu | Met | 368 |
|    | 369 | Ser | Ala | Thr | Gly | Arg | Glu | Met | Ala | Asn | Ile | Gly | Gln | Gln | Ser | Asn | Gln | 384 |
| 10 | 385 | Leu | Pro | Ser | Pro | Val | Ile | Pro | Val | Pro | Thr | Asp | Asp | Ser | Thr | Tyr | Thr | 400 |
|    | 401 | Asn | Tyr | Leu | Arg | Thr | Tyr | Thr | Tyr | Asp | Arg | Gly | Gly | Asn | Leu | Val | Gln | 416 |
| 15 | 417 | Ile | Arg | His | Ser | Ser | Pro | Ala | Thr | Gln | Asn | Ser | Tyr | Thr | Thr | Asp | Ile | 432 |
|    | 433 | Thr | Val | Ser | Ser | Arg | Ser | Asn | Arg | Ala | Val | Leu | Ser | Thr | Leu | Thr | Thr | 448 |
|    | 449 | Asp | Pro | Thr | Arg | Val | Asp | Ala | Leu | Phe | Asp | Ser | Gly | Gly | His | Gln | Lys | 464 |
| 20 | 465 | Met | Leu | Ile | Pro | Gly | Gln | Asn | Leu | Asp | Trp | Asn | Ile | Arg | Gly | Glu | Leu | 480 |
|    | 481 | Gln | Arg | Val | Thr | Pro | Val | Ser | Arg | Glu | Asn | Ser | Ser | Asp | Ser | Glu | Trp | 496 |
| 25 | 497 | Tyr | Arg | Tyr | Ser | Ser | Asp | Gly | Met | Arg | Leu | Leu | Lys | Val | Ser | Glu | Gln | 512 |
|    | 513 | Gln | Thr | Gly | Asn | Ser | Thr | Gln | Val | Gln | Arg | Val | Thr | Tyr | Leu | Pro | Gly | 528 |
|    | 529 | Leu | Glu | Leu | Arg | Thr | Thr | Gly | Val | Ala | Asp | Lys | Thr | Thr | Glu | Asp | Leu | 544 |
| 30 | 545 | Gln | Val | Ile | Thr | Val | Gly | Glu | Ala | Gly | Arg | Ala | Gln | Val | Arg | Val | Leu | 560 |
|    | 561 | His | Trp | Glu | Ser | Gly | Lys | Pro | Thr | Asp | Ile | Asp | Asn | Asn | Gln | Val | Arg | 576 |
| 35 | 577 | Tyr | Ser | Tyr | Asp | Asn | Leu | Leu | Gly | Ser | Ser | Gln | Leu | Glu | Leu | Asp | Ser | 592 |
|    | 593 | Glu | Gly | Gln | Ile | Leu | Ser | Gln | Glu | Glu | Tyr | Tyr | Pro | Tyr | Gly | Gly | Thr | 608 |
|    | 609 | Ala | Ile | Trp | Ala | Ala | Arg | Asn | Gln | Thr | Glu | Ala | Ser | Tyr | Lys | Phe | Ile | 624 |
| 40 | 625 | Arg | Tyr | Ser | Gly | Lys | Glu | Arg | Asp | Ala | Thr | Gly | Leu | Tyr | Tyr | Tyr | Gly | 640 |
|    | 641 | Tyr | Arg | Tyr | Tyr | Gln | Pro | Trp | Val | Gly | Arg | Trp | Leu | Ser | Ala | Asp | Pro | 656 |
| 45 | 657 | Ala | Gly | Thr | Val | Asp | Gly | Leu | Asn | Leu | Tyr | Arg | Met | Val | Arg | Asn | Asn | 672 |
|    | 673 | Pro | Ile | Thr | Leu | Thr | Asp | His | Asp | Gly | Leu | Ala | Pro | Ser | Pro | Asn | Arg | 688 |
|    | 689 | Asn | Arg | Asn | Thr | Phe | Trp | Phe | Ala | Ser | Phe | Leu | Phe | Arg | Lys | Pro | Asp | 704 |
| 50 | 705 | Glu | Gly | Met | Ser | Ala | Ser | Met | Arg | Arg | Gly | Gln | Lys | Ile | Gly | Arg | Ala | 720 |
|    | 721 | Ile | Ala | Gly | Gly | Ile | Ala | Ile | Gly | Gly | Leu | Ala | Ala | Thr | Ile | Ala | Ala | 736 |
| 55 | 737 | Thr | Ala | Gly | Ala | Ala | Ile | Pro | Val | Ile | Leu | Gly | Val | Ala | Ala | Val | Gly | 752 |
|    | 753 | Ala | Gly | Ile | Gly | Ala | Leu | Met | Gly | Tyr | Asn | Val | Gly | Ser | Leu | Leu | Glu | 768 |
|    | 769 | Lys | Gly | Gly | Ala | Leu | Leu | Ala | Arg | Leu | Val | Gln | Gly | Lys | Ser | Thr | Leu | 784 |
| 60 | 785 | Val | Gln | Ser | Ala | Ala | Gly | Ala | Ala | Ala | Gly | Ala | Ser | Ser | Ala | Ala | Ala | 800 |
|    | 801 | Tyr | Gly | Ala | Arg | Ala | Gln | Gly | Val | Gly | Val | Ala | Ser | Ala | Ala | Gly | Ala | 816 |
| 65 | 817 | Val | Thr | Gly | Ala | Val | Gly | Ser | Trp | Ile | Asn | Asn | Ala | Asp | Arg | Gly | Ile | 832 |
|    | 833 | Gly | Gly | Ala | Ile | Gly | Ala | Gly | Ser | Ala | Val | Gly | Thr | Ile | Asp | Thr | Met | 848 |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
|    | 849  | Leu Gly Thr Ala Ser Thr Leu Thr His Glu Val Gly Ala Ala Ala Gly | 864  |
| 5  | 865  | Gly Ala Ala Gly Gly Met Ile Thr Gly Thr Gln Gly Ser Thr Arg Ala | 880  |
|    | 881  | Gly Ile His Ala Gly Ile Gly Thr Tyr Tyr Gly Ser Trp Ile Gly Phe | 896  |
|    | 897  | Gly Leu Asp Val Ala Ser Asn Pro Ala Gly His Leu Ala Asn Tyr Ala | 912  |
| 10 | 913  | Val Gly Tyr Ala Ala Gly Leu Gly Ala Glu Met Ala Val Asn Arg Ile | 928  |
|    | 929  | Met Gly Gly Gly Phe Leu Ser Arg Leu Leu Gly Arg Val Val Ser Pro | 944  |
| 15 | 945  | Tyr Ala Ala Gly Leu Ala Arg Gln Leu Val His Phe Ser Val Ala Arg | 960  |
|    | 961  | Pro Val Phe Glu Pro Ile Phe Ser Val Leu Gly Gly Leu Val Gly Gly | 976  |
|    | 977  | Ile Gly Thr Gly Leu His Arg Val Met Gly Arg Glu Ser Trp Ile Ser | 992  |
| 20 | 993  | Arg Ala Leu Ser Ala Ala Gly Ser Gly Ile Asp His Val Ala Gly Met | 1008 |
|    | 1009 | Ile Gly Asn Gln Ile Arg Gly Arg Val Leu Thr Thr Thr Gly Ile Ala | 1024 |
| 25 | 1025 | Asn Ala Ile Asp Tyr Gly Thr Ser Ala Val Gly Ala Ala Arg Arg Val | 1040 |
|    | 1041 | Phe Ser Leu 1043                                                |      |

30

We claim:

1. A composition, comprising an effective amount of a  
Photorhabdus protein toxin that has functional activity against  
5 an insect.

2. The composition of Claim 1, wherein the *Photorhabdus*  
toxin is produced by a purified culture of *Photorhabdus*, a  
transgenic plant, Baculovirus, or heterologous microbial host.  
10

3. The composition of Claim 2, wherein the *Photorhabdus*  
toxin produced by a purified culture of *Photorhabdus luminescens*.

4. The composition of Claim 2, wherein the toxin is  
15 produced from a purified culture of *Photorhabdus luminescens*  
strain designated ATCC 55397.

5. The composition of Claim 2, wherein the toxin is  
produced by a purified culture of *Photorhabdus luminescens* strain  
20 designated W-14.

6. The composition of Claim 1, wherein the toxin is  
produced by a purified culture of *Photorhabdus* strain designated  
WX-1, WX-2, WX-3, WX-4, WX-5, WX6, WX-7, WX-8, WX-9, WX-10, WX-  
25 11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1, W30, WIR, ATCC#  
43948, ATCC# 43949, ATCC# 43950, ATCC# 43951, or ATCC# 43952.

7. The composition of Claim 2, wherein the toxin is  
produced from a purified culture of *Photorhabdus luminescens*  
30 strain designated WX-1, WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8,  
WX-9, WX-10, WX-11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1,  
W30, WIR, ATCC# 43948, ATCC# 43949, ATCC# 43950, ATCC# 43951, or  
ATCC# 43952.

8. The composition of Claim 1, wherein the toxin is  
35 respresented by amino acid sequence is SEQ ID NO:12.

9. The composition of Claim 6, wherein the composition is a  
mixture of one or more toxins produced from purified cultures of  
40 *Photorhabdus*.

10. The composition of Claim 1 or 6, wherein the insect is of the order *Lepidoptera*, *Coleoptera*, *Hymenoptera*, *Diptera*, *Dictyoptera*, *Acarina* or *Homoptera*.

5

11. The composition of Claim 1 or 6, wherein the insect species is from order *Coleoptera* and is Southern Corn Rootworm, Western Corn Rootworm, Colorado Potato Beetle, Mealworm, Boll Weevil or Turf Grub.

10

12. The composition of Claim 1 or 6, wherein the insect species is from order *Lepidoptera* and is Beet Armyworm, Black Cutworm, Cabbage Looper, Codling Moth, Corn Earworm, European Corn Borer, Tobacco Hornworm, or Tobacco Budworm.

15

13. The composition of Claim 1 or 6, wherein the toxin is formulated as a sprayable insecticide.

20

14. The composition of Claim 1 or Claim 6, wherein the toxin is formulated as a bait matrix and delivered in an above ground or below ground bait station.

25

15. A method of controlling an insect, comprising orally delivering to an insect an effective amount of a protein toxin that has functional activity against an insect, wherein the protein is produced by a purified bacterial culture of the genus *Photobacterium*.

30

16. The method of Claim 15, wherein the bacterium is a purified culture of *Photobacterium luminescens*.

35

17. The method of Claim 15, wherein the toxin is produced from a purified culture of *Photobacterium luminescens* strain designated ATCC 55397.

18. The method of Claim 16, wherein the toxin is produced from a purified culture of *Photobacterium luminescens* strain designated W-14.



19. The method of Claim 15, wherein the toxin is produced from a purified culture of *Photobhabdus* strains designated WX-1, WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-9, WX-10, WX-11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1, W30, WIR, ATCC# 43948, ATCC# 43949, ATCC# ATCC# 43950, ATCC# 43951, or ATCC# 43952.

20. The method of Claim 15, wherein the toxin is produced from a purified culture of *Photobhabdus luminescens* strains designated WX-1, WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-9, WX-10, WX-11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1, W30, WIR, ATCC# 43948, ATCC# 43949, ATCC# ATCC# 43950, ATCC# 43951, or ATCC# 43952.

21. The method of Claim 19, wherein a mixture of one or more toxins is produced from a purified culture of *Photobhabdus* and said toxins are orally delivered to an insect.

22. The method of Claim 15, wherein the toxin is produced by a prokaryotic host transformed with a gene encoding the toxin.

23. The method of Claim 15, wherein the toxin is produced by a eukaryotic host transformed with a gene encoding the toxin.

24. The method of Claim 23, wherein the eukaryotic host is baculovirus.

25. The method of Claim 15 or 19, wherein the insect is of the order *Lepidoptera*, *Coleoptera*, *Hymenoptera*, *Diptera*, *Dictyoptera*, *Acarina* or *Homoptera*.

26. The method of Claim 15 or 19, wherein the insect species is from order *Coleoptera* and is Southern Corn Rootworm, Western Corn Rootworm, Colorado Potato Beetle, Mealworm, Boll Weevil or Turf Grub.

27. The method of Claim 15 or 19, wherein the insect species is from order *Lepidoptera* and is Beet Armyworm, Black Cutworm, Cabbage Looper, Codling Moth, Corn Earworm, European Corn Borer, Tobacco Hornworm, or Tobacco Budworm.

28. The method of Claim 15 or 19, wherein the toxin is formulated as a sprayable insecticide.

29. The method of Claim 15 or Claim 19, wherein the toxin is formulated as a bait matrix and delivered in an above ground or below ground bait station.

30. A method of isolating a gene coding for a protein subunit, comprising the steps of: constructing at least one RNA or DNA oligonucleotide molecule that corresponds to at least a part of a DNA coding region of an amino acid sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO: 13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:43, wherein the nucleotide molecule is used to isolate genetic material from *Photobhabdus* or *Photobhabdus luminescens*.

31. A method for expressing a protein produced by a purified bacterial culture of the genus *Photobhabdus* in a prokaryotic or eukaryotic host in an effective amount so that the protein has functional activity against an insect, wherein the method comprises: constructing a chimeric DNA construct having 5' to 3' a promoter, a DNA sequence encoding a protein, a transcription terminator, and then transferring the chimeric DNA construct into the host.

32. The method of Claim 31, wherein the protein has functional activity against insects selected from a group consisting of Coleoptera, Lepidoptera, Diptera, Homoptera, Hymenoptera, Dictyoptera, and Acarina.

33. The method of Claim 31, wherein the protein encoded by the DNA sequence has an N-terminal amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ

ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:43.

5

34. The method of Claim 31, wherein the protein encoded by the DNA sequence includes the amino acid sequence selected from the group consisting of SEQ ID NO:12, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59 and SEQ ID NO:61.

35. A chimeric DNA construct, adapted for expression in a prokaryotic or eukaryotic host comprising, 5' to 3' a transcriptional promoter active in the host; a DNA sequence encoding a *Photorhabdus* protein that has functional activity against an insect; and a transcriptional terminator.

36. A chimeric DNA construct of Claim 35, wherein the protein encoded by the DNA sequence has an N-terminal amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:43.

37. The chimeric DNA construct of Claim 35, wherein the protein encoded by the DNA sequence has an amino acid sequence selected from the group consisting of SEQ ID NO:12, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, and SEQ ID NO:61.

35

38. The chimeric DNA construct of Claim 35, wherein the DNA sequence encoding the *Photorhabdus luminescens* protein is selected from the group comprising SEQ ID NO:11, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID

NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54,  
SEQ ID NO:56, SEQ ID NO: 58, and SEQ ID NO:60.

39. The chimeric DNA construct of Claim 35, wherein the  
5 host is baculovirus.

40. An isolated and substantially purified preparation  
comprising, a DNA molecule capable of encoding an effective  
amount of a protein that is produced by a bacterium of the genus  
10 *Photorhabdus* and that has functional activity against an insect.

41. The preparation of Claim 40, wherein the bacterium is  
*Photorhabdus luminescens*.

15 42. A purified preparation comprising, a protein produced  
by *Photorhabdus* or *Photorhabdus luminescens* having an N-terminal  
amino acid sequence selected from the group consisting of SEQ ID  
NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID  
NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID  
20 NO: 13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17,  
SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID  
NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:38, SEQ ID NO:39,  
SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:43.

25 43. A purified protein preparation comprising, a protein  
that has an N-terminal amino acid sequence selected from the  
group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID  
NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID  
NO:9, and SEQ ID NO:10, SEQ ID NO: 13, SEQ ID NO:14, SEQ ID NO:15,  
30 SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID  
NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24,  
SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID  
NO:42, and SEQ ID NO:43.

35 44. A purified protein preparation comprising, a protein  
selected from the group of SEQ ID NO:12, SEQ ID NO:26, SEQ ID  
NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:35,  
SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID  
NO:55, SEQ ID NO:57, SEQ ID NO:59, and SEQ ID NO:61.

40

45. A purified DNA preparation comprising, a DNA sequence selected from the group consisting of SEQ ID NO:11, SEQ ID NO:35, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, 5 SEQ ID NO:56, SEQ ID NO:58 and SEQ ID NO:60, wherein the DNA sequence is isolated from its native host.

46. A purified protein preparation comprising, a *Photorhabdus luminescens* protein with at least one subunit having 10 an approximate molecular weight between 18 kDa to about 230 kDa; between about 160 kDa to about 230 kDa; 100 kDa to 160 kDa; about 80 kDa to about 100 kDa; or about 50 kDa to about 80 kDa.

47. A purified protein preparation comprising, a 15 *Photorhabdus luminescens* protein with at least one subunit having an approximate molecular weight of about 280 kDa.

48. A substantially pure microorganism culture comprising, ATCC 55397. 20

49. The culture of Claim 48, wherein the culture is a derivative of ATCC 55397 that produces a protein toxin that has functional activity against an insect.

50. A substantially pure microorganism culture comprising, H9. 25

51. A substantially pure microorganism culture comprising, Hb. 30

52. A substantially pure microorganism culture comprising, Hm.

53. A substantially pure microorganism culture comprising, HP88. 35

54. A substantially pure microorganism culture comprising, NC-1.

55. A substantially pure microorganism culture comprising, 40

W30.

56. A substantially pure microorganism culture comprising,  
WIR.

5

57. A transgenic plant comprising in its genome, a chimeric artificial gene construction imbuing the plant with an ability to express an effective amount of a *Photorhabdus* protein that has functional activity against an insect.

10

58. The transgenic plant of Claim 57, wherein the plant is transformed using acceleration of genetic material coated onto microparticles directly into cells, *Agrobacteria*, whiskers, or electroporation techniques

15

59. The transgenic plant of Claim 57, wherein the selectable marker is selected from the group consisting of kanamycin, neomycin, glyphosate, hygromycin, methotrexate, phosphinothricin (bialophos), chlorosulfuron, bromoxynil, dalapon and the like.

20

60. The transgenic plant of Claim 57, wherein the promoter is selected from the group consisting of octopine synthase, nopaline synthase, mannopine synthase, 35S, 19S, ribulose-1,6-bisphosphate (RUBP) carboxylase small subunit (ssu), beta-conglycinin, phaseolin, alcohol dehydrogenase (ADH), heat-shock, ubiquitin, zein, oleosin, napin, or acyl carrier protein (ACP).

25

61. The transgenic plant of Claim 57, wherein embryogenic tissue, callus tissue type I or II, hypocotyl, meristem, or plant tissue during dedifferentiation is used in preparing the transgenic plant.

30

62. The transgenic plant of Claim 57, wherein the chimeric gene is a DNA sequence which encodes a *Photorhabdus* protein that has functional activity against an insect and at least one codon of the gene has been modified so that the codon is a plant preferred codon.

35

63. A method of controlling an insect comprising orally delivering to an insect an effective amount of a protein toxin, wherein the protein is produced by a transgenic plant, which said insect feeds.

5

64. A composition of matter, comprising a purified DNA sequence from a purified bacterial culture from the genus *Photorhabdus*.

1 ATG CAG GAT TGT COG GAA GTA TCG ATT ACA ACG CTG TCA CTT COC AAA GGT GGC GGT  
 TAC GTC CTA ACA GGC CTT CAT AGC TAA TGT TGC GAC AGT GAA GGG TTT CCA COG CCA  
 10 Met Gln Asp Cys Pro Glu Val Ser Ile Thr Thr Leu Ser Leu Pro Lys Gly Gly Gly  
 P2Psh  
 58 GCT ATC AAT GGC ATG CGA GAA GCA CTG AAT GCT GGC GGC OCT GAT GGA ATG GGC TOC  
 CGA TAG TTA OCG TAC OCT CTT CGT GAC TTA CGA CCG OCG GGA CTA OCT TAC CCG AGG  
 20 Ala Ile Asn Gly Met Gly Glu Ala Leu Asn Ala Ala Gly Pro Asp Gly Met Ala Ser  
 115 CTA TCT CTG CCA TTA OOC CTT TCG ACC GGC AGA GGG ACG GCT CCT GGA TTA TCG CTG  
 GAT AGA GAC GGT AAT GGG GAA AGC TGG CCG TCT OOC TGC CGA GGA OCT AAT AGC GAC  
 39 Leu Ser Leu Pro Leu Pro Leu Ser Thr Gly Arg Gly Thr Ala Pro Gly Leu Ser Leu  
 172 ATT TAC AGC AAC AGT GCA GGT AAT GGG OCT TTC GGC ATC GGC TGG CAA TGC GGT GTT  
 TAA ATG TCG TTG TCA OGT CCA TTA OOC GGA AAG OOG TAG CCG ACC GTT ACG CCA CAA  
 58 Ile Tyr Ser Asn Ser Ala Gly Asn Gly Pro Phe Gly Ile Gly Trp Gln Cys Gly Val  
 229 ATG TOC ATT AGC CGA OGC ACC CAA CAT GGC CTT CAA CAT TGA CGA OGT  
 TAC AGG TAA TCG GCT GCG TGG GTT GTA CCG GAA GTT GTA ACT GCT GCA  
 77 Met Ser Ile Ser Arg Arg Thr Gln His Gly Leu Gln His ... Arg Arg  
 P2.3.5R

FIG. 1



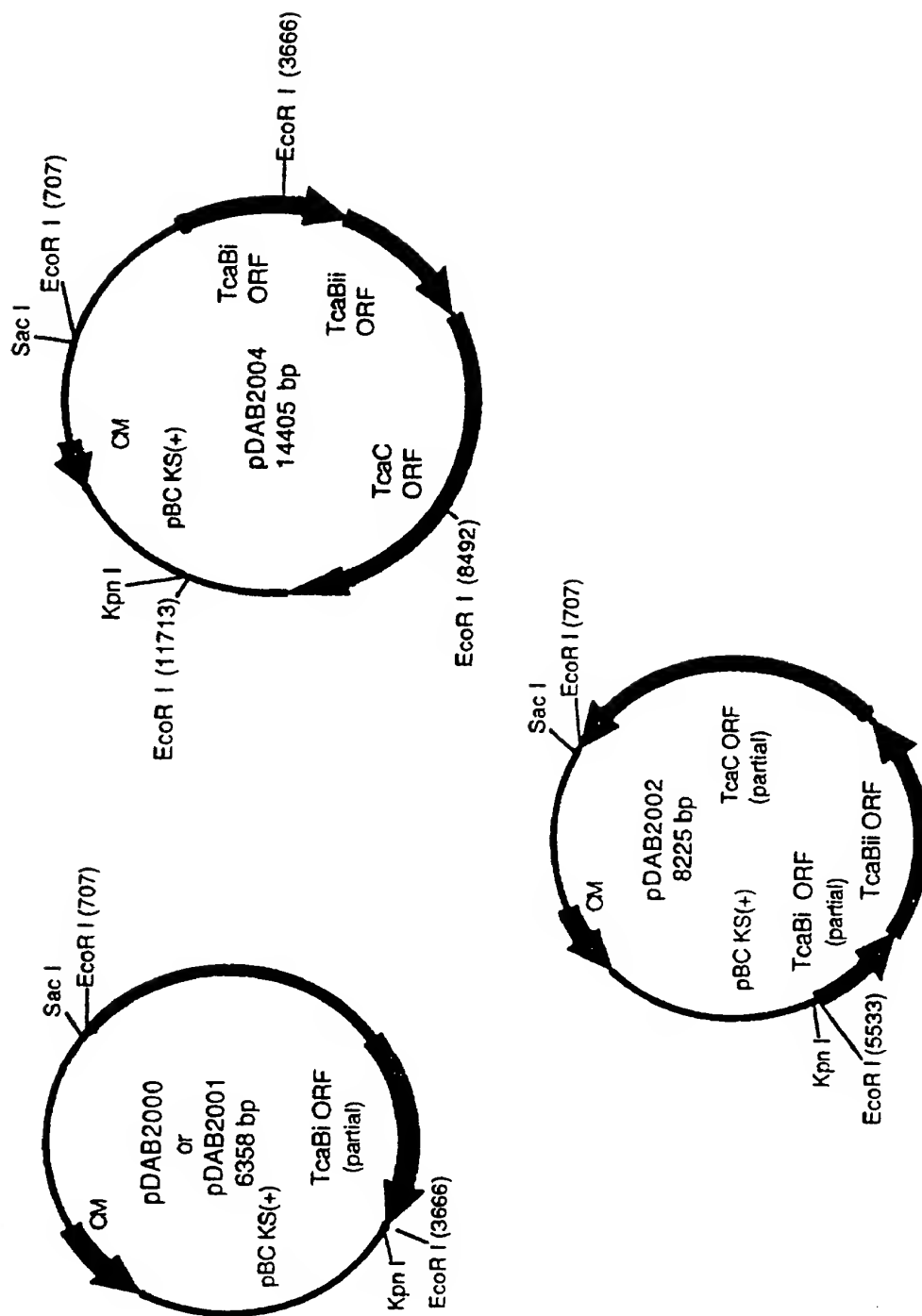


FIG. 2

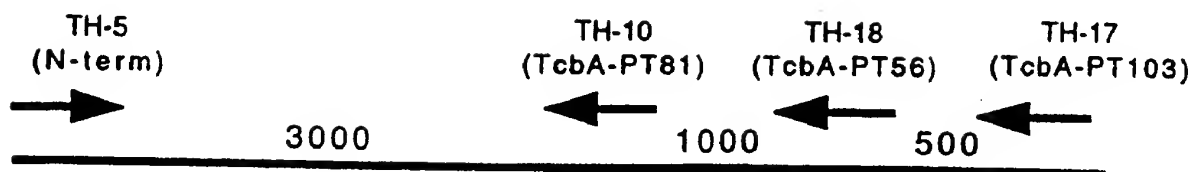


FIG. 3

|        |            |            |             |            |             |
|--------|------------|------------|-------------|------------|-------------|
| TcbA   | 1740       | 1750       | 1760        | 1770       | 1780        |
|        | SSAQALKNDS | EPMDFSGANA | LYFWELFYIT  | PMMMAHRLLO | EQNFDAANHW  |
| TcaBi  | gS         | nPvDFSgpyg | iYlWEiFfhi  | PflvtvRmqt | EQryedAdtW> |
|        | ^^         | ^^^v^      | ^^^v^       | ^^^v^      | ^^^v^       |
| TcbA   | 1790       | 1800       | 1810        | 1820       | 1830        |
|        | FRYVWSPSGY | IVDGKIAIYH | WNVRLPLEEDT | SWNAQQLDST | DPDAVAQDDP  |
| TcaBi  | rdangql    | ImDGskprY- | WNVmPLqldT  | aWdttQpatT | DPDviAmaDP> |
|        | 490        | 510        | 520         | 530        |             |
|        | ^^^v^      | ^^^v^      | ^^^v^       | ^^^v^      | ^^^v^       |
| TcbA   | 1840       | 1850       | 1860        | 1870       | 1880        |
|        | MHYKVATFMA | TLDLLMARGD | AAYRQLERDT  | LAEAKMWYTO | ALNLLGDEPO  |
| TcaBi  | MHYKlAiflh | TLDLLiARGD | sAYRQLERDT  | LvEAKMyYiQ | AqQLLGprPd> |
|        | 540        | 550        | 560         | 570        | 580         |
|        | ^^^v^      | ^^^v^      | ^^^v^       | ^^^v^      | ^^^v^       |
| TcbA   | 1890       | 1900       | 1910        | 1920       | 1930        |
|        | VMLSTTWANP | TLGNAASKTT | QOVRQOVLTO  | LRLNSRVKTP | LLGTANSLTA  |
| TcaBi  | ihhtnTWpNP | TLsk>      |             |            |             |
|        | 600        |            |             |            |             |
|        | ^^^v^      | ^^^v^      |             |            |             |
| TcbA   | 1940       | 1950       | 1960        | 1970       | 1980        |
|        | LFLPQENSKL | KGyWRTLAQR | MFNLRHNLST  | DGQPLSLPLY | AKPADPKALL  |
| TcaBii | FLPpyNdvL  | lGYWdkLeIR | lyNLRHNLST  | DGQPLnLPLY | AtPvDPKtLq> |
|        | 20         | 30         | 40          | 50         | 60          |
|        | ^^^v^      | ^^^v^      | ^^^v^       | ^^^v^      | ^^^v^       |
| TcbA   | 1990       | 2000       | 2010        | 2020       | 2030        |
|        | SAAVSASQGG | ADLPKAPLTI | HRFPQMLEGA  | RGLVNQLIQF | GSSLLGYSER  |
| TcaBii | rqqaggdgtG | sspaggqgsv | qRyPlLvErA  | RsaVslLtQF | GnSLqtLlEh> |
|        | 70         | 80         | 90          | 100        | 110         |
|        | ^^^v^      | ^^^v^      | ^^^v^       | ^^^v^      | ^^^v^       |
| TcbA   | 2040       | 2050       | 2060        | 2070       | 2080        |
|        | QDAEAMSQLL | QTQASELILT | SIRMQDNQLA  | ELDSEKTALQ | VSLAGVQQRf  |
| TcaBii | QDnEkMtill | QTQqeaIlkh | ghdiQqNnLk  | gLqhsLtALQ | aSrdGdtlRq> |
|        | 120        | 130        | 140         | 150        | 160         |
|        | ^^^v^      | ^^^v^      | ^^^v^       | ^^^v^      | ^^^v^       |

FIG. 4A

FIG. 4B

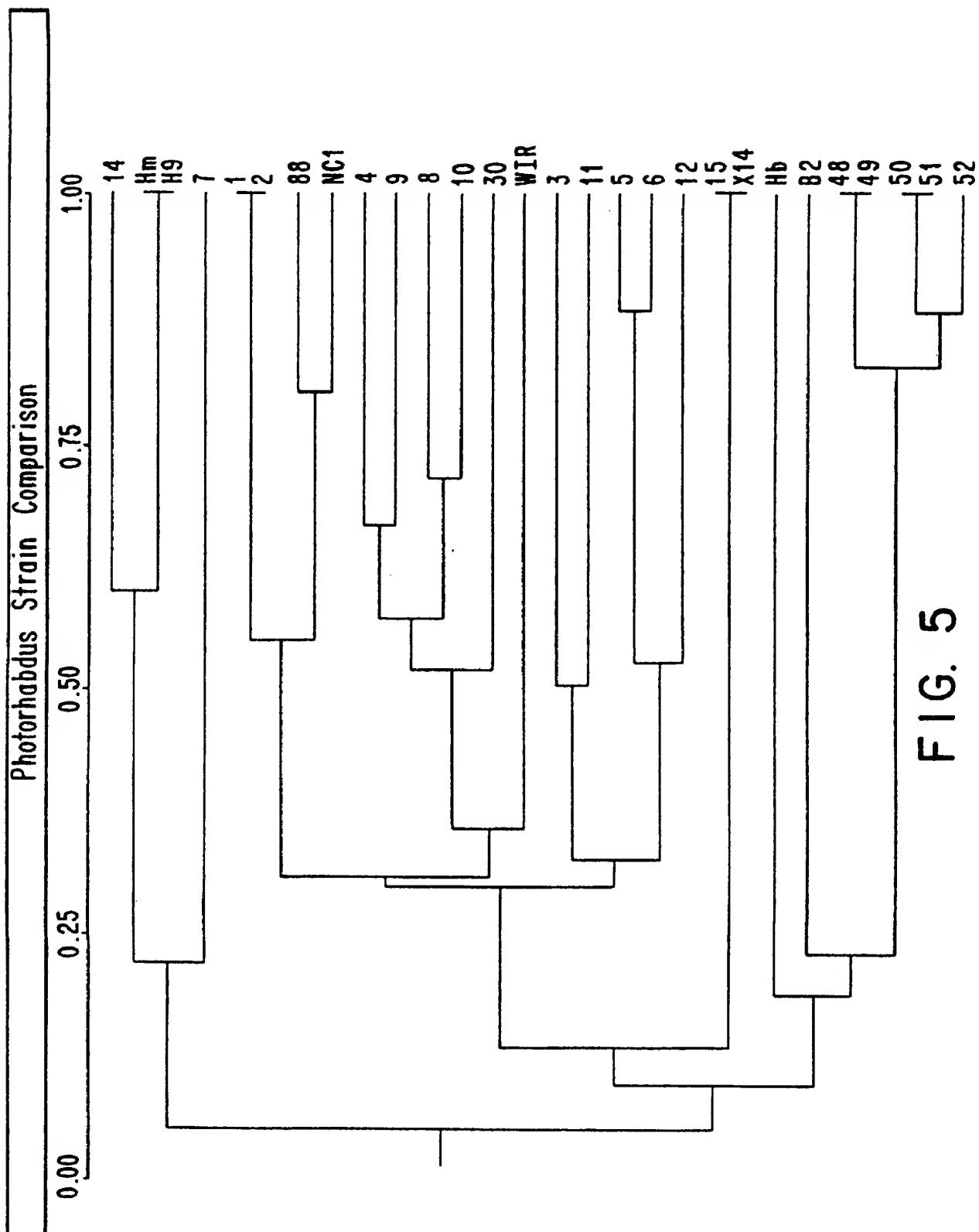


FIG. 5

7 / 8

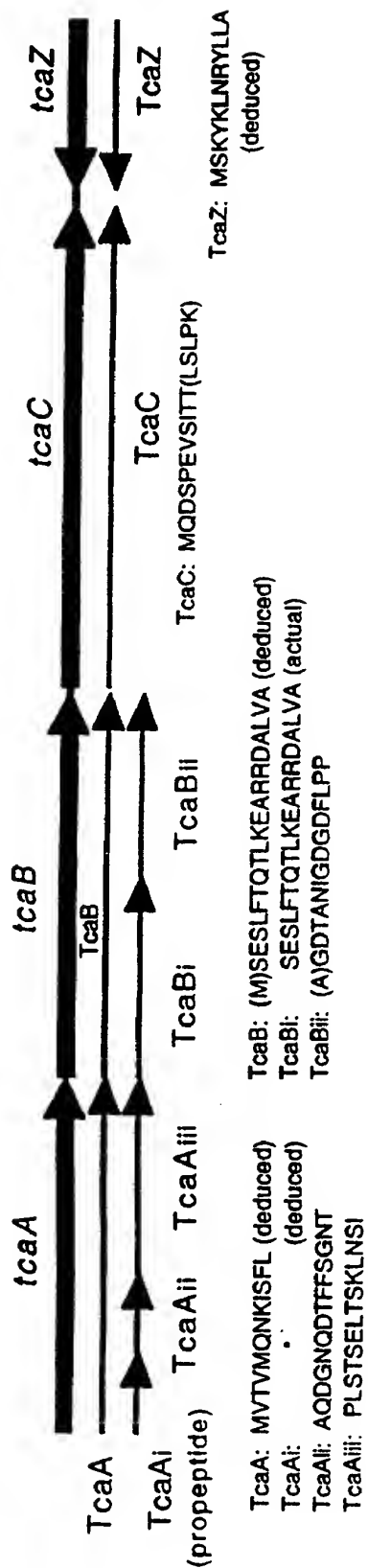
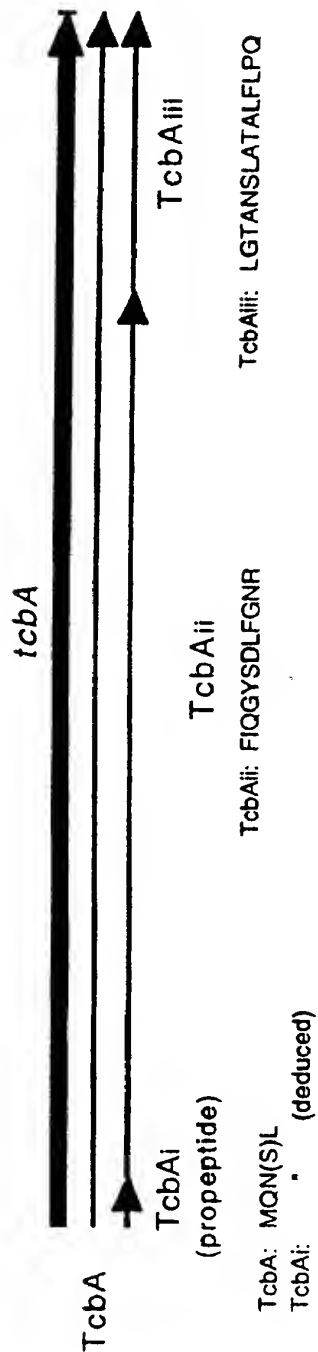
Operon *tca*Locus *tcb*

FIG. 6A

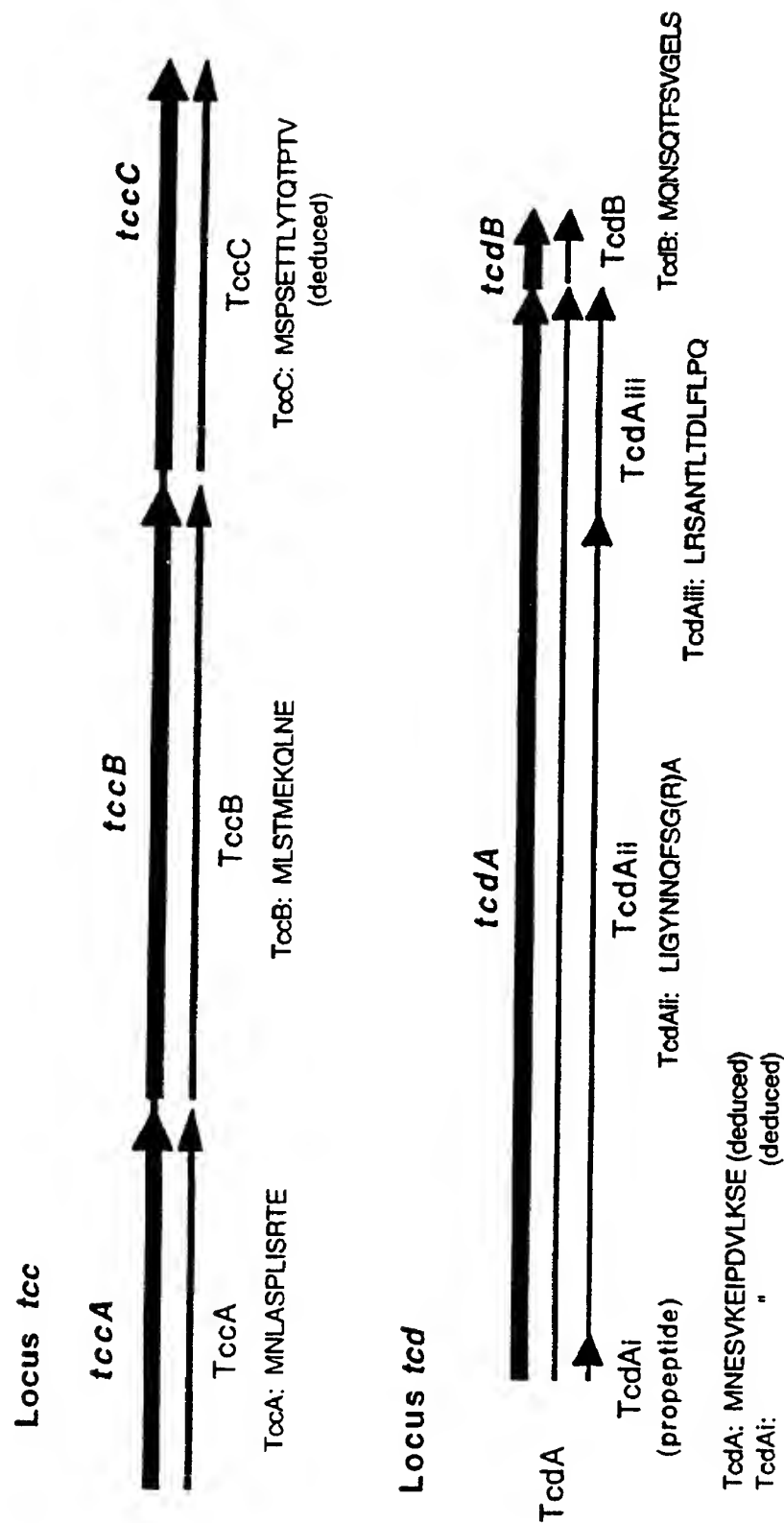


FIG. 6B

## INTERNATIONAL SEARCH REPORT

International application N.  
PCT/US96/18003**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : Please See Extra Sheet.

US CL : 536/23.7, 24.1; 435/172.3, 240.4, 320.1; 800/205; 47/58

According to International Patent Classification (IPC) r to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.7, 24.1; 435/172.3, 240.4, 320.1; 800/205; 47/58

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CABA, CAPLUS, MEDLINE, GENBANK, BIOSIS

search terms: photorhabdus, xenorhabdus, luminescens, insecticide, nematode, lepidoptera

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                      | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | CLARKE et al. Virulence Mechanisms of <i>Photorhabdus</i> sp. Strain K122 toward Wax Moth Larvae. Journal of Invertebrate Pathology. 1995, Vol. 66, pages 149-155, see entire document. | 1-64                  |
| Y         | US 5,039,523 A (PAYNE ET AL.) 13 August 1991, columns 1-10.                                                                                                                             | 1-64                  |
| Y         | US 5,254,799 A (DE GREVE ET AL.) 19 October 1993, columns 1-14.                                                                                                                         | 1-64                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Special categories of cited documents:                                                                                                                                  |                                                                                                                                                                                                                                                  |
| *A* document defining the general state of the art which is not considered to be of particular relevance                                                                | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention                                              |
| *E* earlier document published on or after the international filing date                                                                                                | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone                                                                     |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *O* document referring to an oral disclosure, use, exhibition or other means                                                                                            | *Z* document member of the same patent family                                                                                                                                                                                                    |
| *P* document published prior to the international filing date but later than the priority date claimed                                                                  |                                                                                                                                                                                                                                                  |

Date of the actual completion of the international search

23 DECEMBER 1996

Date of mailing of the international search report

28 JAN 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile N. (703) 305-3230

Authorized officer

THOMAS HAAS

Telephone N. (703) 308-0196



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US96/18003

## A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

C12N 5/14, 15/00, 15/05, 15/09, 15/29, 15/31, 15/64, 15/82; A01G 13/00; A01H 1/00